Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

03078772.5

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03078772.5
Demande no:

Anmeldetag:
Date of filing: 01.12.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

ViroNovative B.V.
Burgemeester Oudlaan 50
3062 PA Rotterdam
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Novel atypical pneumonia-causing virus

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

EP/18.11.03/EP 03078613

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N7/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

Title: Novel atypical pneumonia-causing virus

The invention relates to the field of virology.

Recently, a respiratory illness (atypical pneumonia) was diagnosed in an 8 months old patient that could not be attributed to SARS (Severe Acute Respiratory Syndrome) virus or any other known viral infection. The patient tested negative for influenza, parainfluenza, mumps and RSV and yet the disease was identified to be caused by a virus which closely resembled SARS.

For being able to trace its origin, monitor its epidemiology and prevent possible spreading of the disease, it is of great importance to be able to recognise viral causes of pneumonia in an early stage. Especially, if severe diseases are found to be caused by viruses, it is necessary to detect the identity of the virus as soon as possible, in order to develop diagnostic tools and possibly therapies. The SARS epidemy has shown that it paramount for prevention of spread of the disease to be able to get an early diagnosis in order to timely take effective isolation measures en initiate quarantaine precautions. Only then, world-wide contaminations can be prevented.

Furthermore, identification of the viral cause for the disease enables development of vaccines, which can be used prophylactically to protect people who are risk of being infected. And, finally, knowledge of the viral cause enables to develop therapeutic measures.

Thus, there is great need in developing diagnostic tools and therapies for viral pneumonias in general, and particular to a novel disease-causing infectious agent, especially when this agent appears to be a virus.

The invention provides the nucleotide sequence of an isolated essentially mammalian positive-sense single stranded RNA virus belonging to the Coronaviruses, which is the causative factor for the new disease, hereinafter referred to as EMCR-CoV and the disease being referred to as EMCR-CoV-caused pneumonia. From a phylogenet analysis of the Matrix and Nucleocapsid gene sequences of the virus (Fig. 2a and 2b) it appears that the virus is a distinct member of the group formed by PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine

respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus). Based on amino acid identity matrices, human coronavirus 229E seems to be the closest relative (for all ORFs with the exception of Matrix which appears to be slightly more closely related to PEDV – see Figure 3).

Although phylogenetic analyses provide a convenient method of identifying a virus, several other possibly more straightforward albeit somewhat more coarse methods for identifying said virus or viral proteins or nucleic acids from said virus are herein also provided. As a rule of thumb an EMCR-Coronavirus can be identified by the percentages of homology of the virus, proteins or nucleic acids to be identified in comparison with viral proteins or nucleic acids identified herein by sequence. It is generally known that virus species, especially RNA virus species, often constitute a quasi species wherein a cluster of said viruses displays heterogeneity among its members. Thus it is expected that each isolate may have a somewhat different percentage relationship with the sequences of the isolate as provided herein.

When one wishes to compare a virus isolate with the sequences as listed in figur 1molo, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and determining that said nucleic acid sequence has a percentage nucleic acid identity to the sequences as listed higher than the percentages identified herein for the nucleic acids as identified herein below in comparison with PEDV, 229E, PRCoV, TGE\ CaCoV and FeCoV. Likewise, an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining an amino acid sequence of said virus and determining that said amino acid sequence has a percentage amino acid homology to the sequences as listed which is essentially higher than the percentages provided herein in comparison with PEDV, 229E, PRCoV, TGEV, CaCoV and FeCoV.

With the provision of the sequence information of this EMCR-Coronavirus (EMCR-CoV), the invention provides diagnostic means and methods, prophylactic mea and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease (atypical pneumonia), in particular of mammals, more in particular in humans associated with infection by this virus. In virology, it is most advisory that diagnosis, prophylaxis and/

treatment of a specific viral infection is performed with reagents that are most specifi for said specific virus causing said infection. In this case this means that it is preferre that said diagnosis, prophylaxis and/or treatment of an EMCR-CoV virus infection is performed with reagents that are most specific for EMCR-CoV virus. This by no mean however excludes the possibility that less specific, but sufficiently cross-reactive reagents are used instead, for example because they are more easily available and sufficiently address the task at hand.

The invention for example provides a method for virologically diagnosing an EMCR-CoV infection of an animal, in particular of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a vir isolate or component thereof by reacting said sample with an EMCR-CoV specific nucl acid or antibody according to the invention, and a method for serologically diagnosing EMCR-CoV infection of a mammal comprising determining in a sample of said mamm the presence of an antibody specifically directed against an EMCR-CoV virus or component thereof by reacting said sample with an EMCR-CoV virus-specific proteinaceous molecule or fragment thereof or an antigen according to the invention.

The invention also provides a diagnostic kit for diagnosing an EMCR-CoV infection comprising an EMCR-CoV virus, an EMCR-CoV virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody according to th invention, and preferably a means for detecting said EMCR-CoV virus, EMCR-CoV virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody, said means for example comprising an excitable group such as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To determine whether an as yet unidentified virus component or synthetic analogue there such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as EMCR-CoV-virus-specific, it suffices to analyse the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid, preferably of at least 10, more preferably at least 25, more preferably at least 40 nucleotides or amino acids (respectively), by sequence homology comparison with the provided EMCR-CoV viral sequences and with known non-EMCR-CoV viral sequences (human coronavirus 299E is preferably used) using for example phylogenetic analyses a provided herein. Depending on the degree of relationship with said EMCR-CoV or non-EMCR-CoV viral sequences, the component or synthetic analogue can be identified.

The invention thus provides the nucleotide sequence of a novel etiological agent, an isolated essentially mammalian positive-sense single stranded RNA virus (herein also called EMCR-CoV virus) belonging to the Coronaviridae family, and EMCR-CoV virus-specific components or synthetic analogues thereof.

Coronaviruses were first isolated from chickens in 1937, while the first human coronavirus was propagated *in vitro* by Tyrell and Bonoe in 1965. There are now about 13 species in this family, which infect cattle, pigs, rodents, cats, dogs, birds and man. Coronavirus particles are irregularly shaped, about 60-220 nm in diameter, with an outer envelope bearing distinctive, 'club-shaped' peplomers (about 20 nm long and 10 nm wide at the distal end). This 'crown-like' appearance give the family its name. The envelope carries two glycoproteins: S, the spike glycoprotein which is involved in cell fusion and is a major antigen, and M, the membrane glycoprotein, which is involved in budding and envelope formation. The genome is associated with a basic phosphoprotein, designated N. The genome of coronaviruses, a single stranded positive-sense RNA strand, is typically 27-31 Kb long and contains a 5' methylated cap and a 3' poly-A tail, by which it can directly function as an mRNA in the infected cell. Initially the 5' ORF 1 (about 20 Kb) is translated to produce a viral polymerase, which then produces a full length negative sense strand. This is used as a template to produce mRNA as a 'nested set' of transcripts, all with identical 5' non-translated leader sequence of 72 nucleotides and coincident 3' polyadenylated ends. Each mRNA thus produced is monocistronic, the genes at the 5' end being translated from the longest mRNA and so on. These unusual cytoplasmic structures are produced not by splicing, but by the polymerase during transcription. Between each of the genes there is a repeated intergenic sequence – AACUAAAC – which interacts with the transcriptase plus cellular factors to splice the leader sequence onto the start of each ORF. In some coronaviruses there are about 8 ORFs, coding for the proteins mentioned above, but also for a heamagglutenin esterase (HE), and several other non-structural proteins.

Newly isolated viruses are phylogenetically corresponding to and thus taxonomically corresponding to EMCR-CoV virus when comprising a gene order and/or amino acid sequence and/or nucleotide sequence sufficiently similar to our prototypic EMCR-CoV virus. The highest amino acid sequence identity, between ORFs of EMCR-CoV virus and any of the known other viruses of the same family to date are withhuma coronavirus 299E or Porcine Epidemic Diarrhea Virus (see Figures 3 and 4). The amino acid identities with human coronavirus 229E ranges from 45% (Nucleoprotein) to 81%

(Replicase 1b); interestingly, Replicase 1a has an identity of just 56% contrasting witl Replicase 1b's 81% identity. EMCR CoV has a closer identity with human coronavirus 229E than with any of the known other viruses of the same family to date for all putative ORFs, with the exception of Matrix, which is slightly more closely related to the Matrix ORF of PEDV. Individual proteins or whole virus isolates with, respectivel higher homology than these mentioned maximum values are considered phylogenetically corresponding and thus taxonomically corresponding to EMCR-CoV virus, and generally will be encoded by a nucleic acid sequence structurally corresponding with a sequence as shown in figure 1. Herewith the invention provides virus phylogenetically corresponding to the isolated virus of which the sequences are depicted in figure 1.

It should be noted that, similar to other viruses, a certain degree of variation c be expected to be found between EMCR-CoV-viruses isolated from different sources.

Also, the viral sequence of the EMCR-CoV virus or an isolated EMCR-CoV viru gene as provided herein for example shows less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably les than 65% nucleotide sequence homology or less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably les than 65% amino acid sequence homology with the respective nucleotide or amino acid sequence of the human coronavirus 299E or Porcine Epidemic Diarrhea Virus as for example can be found in Genbank (for example in accession number af304460 (HCoV 299E) or af353511 (PEDV).

Sequence divergence of EMCR-CoV strains around the world may be somewhat higher, in analogy with other coronaviruses.

The term "nucleotide sequence homology" as used herein denotes the presence o homology between two (poly)nucleotides. Polynucleotides have "homologous" sequences if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence. Sequence comparison between two or more polynucleotides generally performed by comparing portions of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window is generally from about 20 to 200 contiguous nucleotides. The "percentage of sequence homology" for polynucleotides, such as 50, 60, 70, 80, 90, 95, 98, 99 or 100 percent sequence homology may be determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide

sequence in the comparison window may include additions or deletions (i.e. gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by: (a) determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions; (b) dividing the number of matched positions by the total number of positions in the window of comparison; and (c) multiplying the result by 100 to yield the percentage of sequence homology. Optimal alignment of sequences for comparison may be conducted by computerized implementations of known algorithms, or by inspection. Readily available sequence comparison and multiple sequence alignment algorithms are, respectively, the Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. 1990. J. Mol. Biol. 215:403; Altschul, S.F. et al. 1997. Nucleic Acid Res. 25:3389-3402) and ClustalW programs both available on the internet. Other suitable programs include GAP, BESTFIT and FASTA in the Wisconsin Genetics Software Package (Genetics Computer Group (GCG), Madison, WI, USA).

As used herein, "substantially complementary" means that two nucleic acid sequences have at least about 65%, preferably about 70%, more preferably about 80%, even more preferably 90%, and most preferably about 98%, sequence complementarity to each other. This means that the primers and probes must exhibit sufficient complementarity to their template and target nucleic acid, respectively, to hybridise under stringent conditions. Therefore, the primer sequences as disclosed in this specification need not reflect the exact sequence of the binding region on the template and degenerate primers can be used. A substantially complementary primer sequence is one that has sufficient sequence complementarity to the amplification template to result in primer binding and second-strand synthesis.

The term "hybrid" refers to a double-stranded nucleic acid molecule, or duplex, formed by hydrogen bonding between complementary nucleotides. The terms "hybridise' or "anneal" refer to the process by which single strands of nucleic acid sequences form double-helical segments through hydrogen bonding between complementary nucleotides

The term "oligonucleotide" refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alky and aryl-phosphate, phosphorothioate), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl

ribosyl, 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains capable of forming stable secondary structures (e.g., stem-and-loop and loop-stem-loop structures).

The term "primer" as used herein refers to an oligonucleotide which is capable annealing to the amplification target allowing a DNA polymerase to attach thereby serving as a point of initiation of DNA synthesis when placed under conditions in whi synthesis of primer extension product which is complementary to a nucleic acid stran is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxy ribonucleotide. The primer must be sufficiently long to prime t synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primer. A "pair of bi-directional primers" as used herein refers to one forwar and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

The term "probe" refers to a single-stranded oligonucleotide sequence that will recognize and form a hydrogen-bonded duplex with a complementary sequence in a target nucleic acid sequence analyte or its cDNA derivative.

The terms "stringency" or "stringent hybridization conditions" refer to hybridization conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimised to maximize specific binding and minimize non-specific binding primer or probe to its target nucleic acid sequence. The terms as used include referenc to conditions under which a probe or primer will hybridise to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridise specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridises to a perfectly matched probe or primer.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na+ ion, typically about 0.01 to 1.0 M Na+ ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or "conditions of reduced stringency" include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 2x SSC at 40°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1x SSC at 60°C. Hybridization procedures are well known in the art and are described in e.g. Ausubel et al, Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994.

The term "antibody" includes reference to antigen binding forms of antibodies (e. g., Fab, F (ab) 2). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i. e., comprising constant and variable regions from different species), humanized antibodies (i. e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e. g., bispecific antibodies).

In short, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of a suitable fragment of the genome of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 1 than it is corresponding to a virus isolate of PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus).

Suitable nucleic acid genome fragments each useful for such phylogenetic tree analyses are for example any of the fragments encoding the Matrix protein or the Nucleocapsid protein as disclosed in Figure 1, leading to the phylogenetic tree analys as disclosed herein in figure 2a or 2b. Other suitable nucleic acid fragments useful fo such phylogenetic tree analyses are for example any of the fragments encoding Replic 1a and 1b, Spike, orf 4a and 4b, and E.

A suitable open reading frame (ORF) useful in phylogenetic analyses comprise the ORF encoding the viral replicase (ORF 1a). When an overall amino acid identity o at least 60%, preferably of at least 70%, more preferably of at least 80%, more prefera of at least 90%, most preferably of at least 95% of the analysed replicase with the replicase having a sequence comprising the amino acids of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the inventic

A suitable open reading frame (ORF) useful in phylogenetic analyses comprise the ORF encoding the viral replicase (ORF 1b). When an overall amino acid identity o at least 82%, more preferably of at least 90%, most preferably of at least 95% of the analysed replicase with the replicase having a sequence comprising the amino acids o Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the Nucleocapsid protein. When an overall amino acid identity of at least 50%, more preferably of at least 60%, more preferably of at least 70 more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed Nucleocapsid protein with the Nucleocapsid protein encoded by a sequence comprising (part of) the sequence F of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the Matrix protein. When an overall amino acid identity at least 60%, more preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed Matrix protein with the Matrix protein encoded by a sequence comprising (part of) the sequen F of Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the spike protein S. When an overall amino acid identity

at least 55%, more preferably of at least 60%, more preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed S-protein encoded by a sequence comprising the sequence of translation 2 of E and translation 1 of the F sequence of the S-protein as depicted in Figure 1 is found, the analysed virus isolate comprises an EMCR-CoV virus isolate according to the invention. The S ORF of the EMCR-CoV virus seems to be located adjacent to the ORF 1ab (coding for the viral replicase), which would discriminate an EMCR-CoV viruses from the bovine coronavirus and the murine hepatitis virus, which have a so-called 2a gene and an HE-gene between the S protein and the viral polymerase.

The invention provides among others an isolated or recombinant nucleic acid or virus-specific functional fragment thereof obtainable from a virus according to the invention. The isolated or recombinant nucleic acids comprises the sequences as given in figure 1 or sequences of homologues which are able to hybridise with those under stringent conditions. In particular, the invention provides primers and/or probes suitable for identifying an EMCR-CoV virus nucleic acid.

Furthermore, the invention provides a vector comprising a nucleic acid according to the invention. To begin with, vectors such as plasmid vectors containing (parts of) the genome of the EMCR-CoV virus, virus vectors containing (parts of) the genome of the EMCR-CoV (for example, but not limited thereto, vaccinia virus, retroviruses, baculovirus), or EMCR-CoV virus containing (parts of) the genome of other viruse or other pathogens are provided.

Also, the invention provides a host cell comprising a nucleic acid or a vector according to the invention. Plasmid or viral vectors containing the replicase components of EMCR-CoV virus are generated in prokaryotic cells for the expression of the components in relevant cell types (bacteria, insect cells, eukaryotic cells). Plasmid or viral vectors containing full-length or partial copies of the EMCR-CoV virus genome will be generated in prokaryotic cells for the expression of viral nucleic acids *in-vitro* or *in-vivo*. The latter vectors may contain other viral sequences for the generation of chimeric viruses or chimeric virus proteins, may lack parts of the viral genome for the generation of replication defective virus, and may contain mutations, deletions or insertions for the generation of attenuated viruses.

Infectious copies of EMCR-CoV virus (being wild type, attenuated, replication-defective or chimeric) can be produced upon co-expression of the polymerase compone according to the state-of-the-art technologies described above.

In addition, eukaryotic cells, transiently or stably expressing one or more full-length or partial EMCR-CoV virus proteins can be used. Such cells can be made by transfection (proteins or nucleic acid vectors), infection (viral vectors) or transduction (viral vectors) and may be useful for complementation of mentioned wild type, attenuated, replication-defective or chimeric viruses.

A chimeric virus may be of particular use for the generation of recombinant vaccines protecting against two or more viruses. For example, it can be envisaged that EMCR-CoV virus vector expressing one or more proteins of a human metapneumoviru or a human metapneumovirus vector expressing one or more proteins of EMCR-CoV virus will protect individuals vaccinated with such vector against both virus infections Such a specific chimeric virus is particularly useful in the invention because it is suspected that co-infection of, for instance, human metapneumovirus frequently occur in coronavirus infected patients. Attenuated and replication-defective viruses may be use for vaccination purposes with live vaccines as has been suggested for other viruses

In a preferred embodiment, the invention provides a proteinaceous molecule or coronavirus-specific viral protein or functional fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derive from any of the genes or genomic fragments derivable from a virus according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as sub-unit vaccines and inhibitory peptides. Particularly useful are the viral replicase protein, the spike protein, the matrix protein, the nucleocapsid or antigenic fragments thereof for inclusion as antigen or subunit immunogen, but inactivated whole virus ca also be used. Particulary useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting EMCR-CoV virus specific antibodies, whether in vivo (e.g. for protective puposes or for providing diagnostic antibodies) or in vitro (e.g. by phage display technology or another technique useful for generating synthetic antibodies).

Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g. (phage) library-derived binding molecules) antibodies that specifically react with an antigen comprising a proteinaceous molecule or EMCR-CoV virus-specific functional fragment thereof according to the invention. Such antibodies are useful in a method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with an antibody as provided herein. This can for example be achieved by using purified or non-purified EMCR-CoV virus or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against EMCR-CoV virus proteins which are encoded by a nucleotide sequence comprising one or more of the sequences disclosed in figure 1.

Other methods for identifying a viral isolate as an EMCR-CoV virus comprise reacting said viral isolate or a component thereof with a virus specific nucleic acid according to the invention.

In this way the invention provides a viral isolate identifiable with a method according to the invention as a mammalian virus taxonomically corresponding to a positive-sense single stranded RNA virus identifiable as likely belonging to the EMCR-CoV virus genus within the family of Coronaviruses.

The method is useful in a method for virologically diagnosing an EMCR-CoV virus infection of a mammal, said method for example comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid or an antibody according to the invention.

Methods of the invention can in principle be performed by using any nucleic acid amplification method, such as the Polymerase Chain Reaction (PCR; Mullis 1987, U.S. Pat. No. 4,683,195, 4,683,202, en 4,800,159) or by using amplification reactions such as Ligase Chain Reaction (LCR; Barany 1991, Proc. Natl. Acad. Sci. USA 88:189-193; EP Appl. No., 320,308), Self-Sustained Sequence Replication (3SR; Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), Strand Displacement Amplification (SDA; U.S. Pat. Nos. 5,270,184, en 5,455,166), Transcriptional Amplification System (TAS; Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), Rolling Circle Amplification (RCA; U.S. Pat. No. 5,871,921), Nucleic Acid Sequence Based Amplification (NASBA), Cleavase Fragment

Length Polymorphism (U.S. Pat. No. 5,719,028), Isothermal and Chimeric Primer-initiated Amplification of Nucleic Acid (ICAN), Ramification-extension Amplification Method (RAM; U.S. Pat. Nos. 5,719,028 and 5,942,391) or other suitable methods for amplification of nucleic acids.

In order to amplify a nucleic acid with a small number of mismatches to one or more the amplification primers, an amplification reaction may be performed under conditi of reduced stringency (e.g. a PCR amplification using an annealing temperature of 38°C, or the presence of 3.5 mM MgCl2). The person skilled in the art will be able to select conditions of suitable stringency.

The primers herein are selected to be "substantially" complementary (i.e. at le 65%, more preferably at least 80% perfectly complementary) to their target regions present on the different strands of each specific sequence to be amplified. It is possib to use primer sequences containing e.g. inositol residues or ambiguous bases or even primers that contain one or more mismatches when compared to the target sequence general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target DNA or RNA oligonucleotide sequences, are considered suitable for u in a method of the present invention. Sequence mismatches are also not critical whe using low stringency hybridization conditions.

The detection of the amplification products can in principle be accomplished b any suitable method known in the art. The detection fragments may be directly stair or labelled with radioactive labels, antibodies, luminescent dyes, fluorescent dyes, or enzyme reagents. Direct DNA stains include for example intercalating dyes such as acridine orange, ethidium bromide, ethidium monoazide or Hoechst dyes. Alternatively, the DNA or RNA fragments may be detected by incorporation of labell dNTP bases into the synthesized fragments. Detection labels which may be associate with nucleotide bases include e.g. fluorescein, cyanine dye or BrdUrd.

When using a probe-based detection system, a suitable detection procedure fo use in the present invention may for example comprise an enzyme immunoassay (EI format (Jacobs et al., 1997, J. Clin. Microbiol. 35, 791-795). For performing a detecti by manner of the EIA procedure, either the forward or the reverse primer used in th amplification reaction may comprise a capturing group, such as a biotin group for immobilization of target DNA PCR amplicons on e.g. a streptavidin coated microtite plate wells for subsequent EIA detection of target DNA -amplicons (see below). The

skilled person will understand that other groups for immobilization of target DNA PCR amplicons in an EIA format may be employed.

Probes useful for the detection of the target DNA as disclosed herein preferably bind only to at least a part of the DNA sequence region as amplified by the DNA amplification procedure. Those of skill in the art can prepare suitable probes for detection based on the nucleotide sequence of the target DNA without undue experimentation as set out herein. Also the complementary nucleotide sequences, whether DNA or RNA or chemically synthesized analogs, of the target DNA may suitably be used as type-specific detection probes in a method of the invention, provided that such a complementary strand is amplified in the amplification reaction employed.

Suitable detection procedures for use herein may for example comprise immobilization of the amplicons and probing the DNA sequences thereof by e.g. southern blotting. Other formats may comprise an EIA format as described above. To facilitate the detection of binding, the specific amplicon detection probes may comprise a label moiety such as a fluorophore, a chromophore, an enzyme or a radio-label, so as to facilitate monitoring of binding of the probes to the reaction product of the amplification reaction. Such labels are well-known to those skilled in the art and include, for example fluorescein isothiocyanate (FITC), β-galactosidase, horseradish peroxidase, streptavidin biotin, digoxigenin, 35S or 125I. Other examples will be apparent to those skilled in the art.

Detection may also be performed by a so called reverse line blot (RLB) assay, such as for instance described by Van den Brule et al. (2002, J. Clin. Microbiol. 40, 779-787). For this purpose RLB probes are preferably synthesized with a 5' amino group for subsequent immobilization on e.g. carboxyl-coated nylon membranes. The advantage of an RLB format is the ease of the system and its speed, thus allowing for high throughput sample processing.

The use of nucleic acid probes for the detection of RNA or DNA fragments is well known in the art. Mostly these procedure comprise the hybridization of the target nucleic acid with the probe followed by post-hybridization washings. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For nucleic acid hybrids, the Tm can be approximated from the equation of Meinkoth and Wahl, Anal. Biochem., 138: 267-284 (1984): $Tm = 81.5\ °C + 16.6\ (\log M) + 0.41\ (\%\ GC) - 0.61\ (\%\ form) - 500/L$; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine

nucleotides in the nucleic acid, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. Tm is reduced by about 1 °C each 1 % of mismatching; thus, the hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with > 90% identity are sought, the Tm can be decreased 10°C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence and its complement at a defined ionic strength and pH. However severely stringent conditions can utilize a hybridization and/or wash at 1,2,3, or 4 °C lower than the thermal melting point (Tm); moderately stringent conditions can utiliz hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the thermal melting point (Tm); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 1 15, or 20 °C lower than the thermal melting point (Tm). Using the equation, hybridization and wash compositions, and desired Tm, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions ar inherently described. If the desired degree of mismatching results in a Tm of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemist and Molecular Biology—Hybridization with Nucleic Acid Probes, Part I, Chapter 2" Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier. New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995).

In another aspect, the invention provides oligonucleotide probes for the generic detection of target RNA or DNA. The detection probes herein are selected to be "substantially" complementary to one of the strands of the double stranded nucleic acid generated by an amplification reaction of the invention. Preferably the probes are substantially complementary to the immobilizable, e.g. biotin labelled, antisense strand of the amplicons generated from the target RNA or DNA.

It is allowable for detection probes of the present invention to contain one or more mismatches to their target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target oligonucleotide sequences are considered suitable for use in a method of the present invention.

Antibodies, both monoclonal and polyclonal, can also be used for detection purpose in the present invention, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. A variety of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that can be used to determine selective binding. Examples of types of immunoassays that can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays that are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

Antibodies can be bound to many different carriers and used to detect the presence of the target molecules. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such using routine experimentation.

The invention also provides a method for serologically diagnosing an EMCR-CoV virus infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against an EMCR-CoV virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof or an antigen according to the invention

Methods and means provided herein are particularly useful in a diagnostic kit for diagnosing an EMCR-CoV virus infection, be it by virological or serological diagnosis. Such kits or assays may for example comprise a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention

Use of a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention is also provided for the product of a pharmaceutical composition, for example for the treatment or prevention of EMC CoV virus infections and/or for the treatment or prevention of atypical pneumonia, in particular in humans. Preferably a peptide comprising part of the amino acid sequenc of the spike protein as depicted in the relevant translations of Figure 1, is used for the preparation of a therapeutic or prophylactic peptide. Also preferably, a protein comprising the amino acid sequence of the spike protein as depicted in the relevant translations of Figure 1, is used for the preparation of a sub-unit vaccine. Furthermor the nucleocapsid of Coronaviruses, as depicted in the translation of Figure 1, is known to be particularly useful for eliciting cell-mediated immunity against Coronaviruses ar can be used for the preparation of a sub-unit vaccine.

Attenuation of the virus can be achieved by established methods developed for this purpose, including but not limited to the use of related viruses of other species, serial passages through laboratory animals or/and tissue/cell cultures, serial passages through cell cultures at temparutes below 37°C (cold-adaption), site directed mutagenesis of molecular clones and exchange of genes or gene fragments between related viruses.

A pharmaceutical composition comprising a virus, a nucleic acid, a proteinaceou molecule or fragment thereof, an antigen and/or an antibody according to the invention can for example be used in a method for the treatment or prevention of an EMCR-CoV virus infection and/or a respiratory illness comprising providing an individual with a pharmaceutical composition according to the invention. This is most useful when said individual comprises a human. Antibodies against EMCR-CoV virus proteins, especiall against the spike protein of EMCR-CoV virus, preferably against the amino acid sequence as depicted in the translation in figure 1, are also useful for prophylactic or therapeutic purposes, as passive vaccines. It is known from other coronaviruses that th spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic vaccination.

The invention also provides method to obtain an antiviral agent useful in the treatment of atypical pneumonia comprising establishing a cell culture or experimental animal comprising a virus according to the invention, treating said culture or animal with an candidate antiviral agent, and determining the effect of said agent on said viru or its infection of said culture or animal. An example of such an antiviral agent

comprises an EMCR-CoV virus-neutralising antibody, or functional component thereof, as provided herein, but antiviral agents of other nature are obtained as well.

The invention also provides use of an antiviral agent according to the invention for the preparation of a pharmaceutical composition, in particular for the preparation of a pharmaceutical composition for the treatment of atypical pneumonia, especifically when caused by an EMCR-CoV virus infection, and provides a pharmaceutical composition comprising an antiviral agent according to the invention, useful in a method for the treatment or prevention of an EMCR-CoV virus infection or atypical pneumonia, said method comprising providing an individual with such a pharmaceutical composition.

The invention also comprises an animal model usable for testing of prophylactic and/or therapeutic methods and/or preparations. It is hypothesized that apes can be infected with the EMCR-CoV virus, thereby showing clinical symptoms, and more importantly, similar tissue morphology as found in humans suffering from atypical pneumonia caused by the EMCR-CoV virus. Subjecting apes to a prophylactic or therapeutic treatment either before or during infection with the virus will have a good and useful predictionary value for application of such a prophylaxis or therapy in human subjects.

The invention is further explained in the Examples without limiting it thereto.

Figure legends

Fig. 1: Nucleotide sequences from parts of the EMCR-CoV virus. Also included are the putative amino acid sequences of polypeptides.

Fig. 2: Phylogenetic relationship for the nucleotide sequences of isolate EMCR-CoV wi its closest relatives genetically. Phylogenetic trees were generated by maximum likelihood analyses using 100 bootstraps and 3 jumbles. The scale representing the number of nucleotide changes is shown for each tree. Figure 1a. Maximum likelihood tree of matrix gene nucleotide sequences. Numbers in trees represent bootstrap value The scale bar roughly reflects 10 % nucleotide differences between related sequences. Figure 1b. Maximum likelihood tree of nucleocapsid gene nucleotide sequences. Numbers in trees represent bootstrap values. The scale bar roughly reflects 10 % nucleotide differences between related sequences.

Fig. 3: Similarity matrices indicating amino acid identity for the putative Replicase 1a Replicase 1b, Replicase 1ab, Spike, Orf E, Matrix and Nucleocapsid proteins 3a-g, respectively), and for the putative Matrix protein and Nucleoprotein (3h and 3i resp.) between the EMCR-CoV virus and closely related coronaviruses. See text for abbreviations.

Figure 4 Alignments with various coronaviruses: 5'untranslated region genomic sequence (a); Putative orf 1a amino acid sequence (b); Putative orf 1b amino acid sequence (c); Putative orf 1ab amino acid sequence (d); Putative Spike amino acid sequence (e); Putative orf 4a amino acid sequence (f); Putative orf 4ab amino acid sequence (g); Putative orf E amino acid sequence (h); Putative Matrix amino acid sequence (i); Putative Nucleoprotein amino acid sequence (j); Putative 3'untranslated genomic sequence (k); See text for abbreviations.

Examples

Specimen collection

Virus was collected from an 8 month old patient suffering from pneumonia using nasal swabs.

Virus isolation and culture

Throat swabs were dipped into a culture of tMK cells and passaged four times. Virus was then in Vero-118 cells. One litre of virus containing cell culture supernatant was harvested, and the virus was pelleted in an ultracentrifuge and the virus pellet was resuspended in1ml PBS.

RNA isolation

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient fractions using a High Pure RNA Isolation kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands).

Sequencing

Purified RNA was sent to BaseClear holding BV (Leiden, The Netherlands) for sequencing.

Phylogenetic analyses

Nucleotide sequences were aligned using Clustal W running under BioEdit version 5.0.9. Maximum likelihood trees were created using the Seqboot and DNA-ML packages of Phylip 5.6 using 100 bootstraps and 3 jumbles. The consensus trees were calculated using the Consense package of phylip 5.6. These consensus trees were used as usertree in DNA-ML to recalculate the branch lengths from the original sequences.

The sequences of EMCR-CoV were compared with those of reference viruses representing each species in the four groups of coronaviruses. These were: human coronavirus 229E (229E), af304460; porcine epidemic diarrhea virus (PEDV) af353511; transmissible gastroenteritis virus (TGEV), aj271965; bovine coronavirus (BoCoV), af220295; murine hepatitis virus (MHV), af201929; avian infectious bronchitis virus (AIBV), m95169, Canine coronavirus (CaCoV), d13096; feline coronavirus (FeCoV),

ay204704; porcine respiratory coronavirus (PRCoV), z24675; human coronavirus OC4 (OC43), m76373, l14643, m933990; porcine haemagglutinating encephalomyelitis viru (HEV), ay078417; rat coronavirus (RtCoV) af 207551) References for the viruses are t numbers of the NCBI catalog (http://www.ncbi.nlm.nih.gov/entrez/).

In general, coronaviruses, such as EMCR-CoV can be isolated and identified accordin to the following protocol:

Specimen collection

In order to find virus isolates nasopharyngeal aspirates, throat and nasal swabs, broncheo alveolar lavages, serum and plasma samples, and stools preferably from mammals such as humans, carnivores (dogs, cats, mustellits, seals etc.), horses, ruminants (cattle, sheep, goats etc.), pigs, rabbits, birds (poultry, ostriches, etc) shoul be examined. From birds cloaca swabs and droppings can be examined as well. Sera should be collected for immunological assays, such as ELISA, molecular-based assays, such as RT-PCR and virus neutralisation assays.

Collected virus specimens may be diluted with 5 ml Dulbecco MEM medium (BioWhittaker, Walkersville, MD) and thoroughly mixed on a vortex mixer for one minute. The suspension is thus centrifuged for ten minutes at 840 x g. The sediment i spread on a multispot slide (Nutacon, Leimuiden, The Netherlands) for immunofluorescence techniques, and the supernatant is used for virus isolation.

Virus isolation

For virus isolation Vero-118 cells or tMK cells (RIVM, Bilthoven, The Netherlands) we cultured in 24 well plates containing glass slides (Costar, Cambridge, UK), with the medium described below supplemented with 10% fetal bovine serum (BioWhittaker, Vervier, Belgium). Before inoculation the plates were washed with PBS and supplied with Eagle's MEM with Hanks' salt (ICN, Costa mesa, CA) supplemented with 0.52/lit gram $NaHCO_3$, 0.025 M Hepes (Biowhittaker), 2 mM L-glutamine (Biowhittaker), 200 units/liter penicilline, 200 µg/liter streptomycine (Biowhittaker), 1gram/liter lactalbumine (Sigma-Aldrich, Zwijndrecht, The Netherlands), 2.0 gram/liter D-glucose (Merck, Amsterdam, The Netherlands), 10 gram/liter peptone (Oxoid, Haarlem, The Netherlands) and 0.02% trypsine (Life Technologies, Bethesda, MD). The plates were inoculated with supernatant of the patient samples, 0,2 ml per well in triplicate, followed by centrifuging at 840x *g* for one hour. After inoculation the plates were

incubated at 37 °C for 1-7 days and cultures were checked daily for CPE. Extensive CPE was generally observed within 5-10 and included detachment of cells from the monolayer..

Virus culture

Sub-confluent monolayers of tMK cells or Vero clone 118 cells in media as described above were inoculated with supernatants of samples that displayed CPE or with samples taken from a patient.

RNA isolation

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient fractions using a High Pure RNA Isolation kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands). RNA can also be isolated following other procedures known in the field (*Current Protocols in Molecular Biology*).

Sequence analysis

Sequence analyses were performed by BaseClear holding BV (Leiden, The Netherlands)

Claims

1. An isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) comprising the sequence of figure 1 or homologues thereof.

2. An isolated positive-sense single stranded RNA virus (EMCR-CoV) belonging t the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 1 than it is corresponding to a virus isolate of PEDV (porcine epidemic diarrhea virus), HCoV-229E (human coronavirus 229E), PRCoV (porcine respiratory coronavirus), TGEV (transmissible gastroenteritis virus), CaCoV (Canine coronavirus) and FeCoV (feline coronavirus).

3. A virus according to claim 1 or 2 wherein said nucleic acid sequence comprises open reading frame (ORF) encoding a viral protein of said virus.

4. A virus according to claim 3 wherein said open reading frame is selected from t group of ORFs encoding the viral replicase, nuclear capsid protein, matrix protein and the spike protein.

5. A virus according to claim 1-4 isolatable from a human with atypical pneumoni

6. An isolated or recombinant nucleic acid or EMCR-CoV virus-specific functional fragment thereof obtainable from a virus according to anyone of claims 1 to 5.

7. A vector comprising a nucleic acid according to claim 6.

8. A host cell comprising a nucleic acid according to claim 6 or a vector according t claim 7.

9. An isolated or recombinant proteinaceous molecule or EMCR-CoV virus-specific functional fragment thereof encoded by a nucleic acid according to claim 6.

10. An antigen comprising a proteinaceous molecule or EMCR-CoV virus-specific functional fragment thereof according to claim 9.

11. An antibody specifically directed against an antigen according to claim 10.

12. A method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with an antibody according to claim 11.

13. A method for identifying a viral isolate as an EMCR-CoV virus comprising reacting said viral isolate or a component thereof with a nucleic acid according to claim 6.

14. A method for virologically diagnosing an EMCR-CoV infection of a mammal comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid according to claim 6 or an antibody according to claim 11.

15. A method for serologically diagnosing an EMCR-CoV infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against an EMCR-CoV virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof according to claim 9 or an antigen according to claim 10.

16. A diagnostic kit for diagnosing an EMCR-CoV infection comprising a virus according to anyone of claims 1 to 5, a nucleic acid according to claim 6, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10 and/or an antibody according to claim 11.

17. Use of a virus according to any one claims 1 to 5, a nucleic acid according to clai 6, a vector according to claim 7, a host cell according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10, or an antibody according to claim 11 for the production of a pharmaceutical composition.

18. Use according to claim 17 for the production of a pharmaceutical composition f the treatment or prevention of an EMCR-CoV virus infection.

19. Use according to claim 17 or 18 for the production of a pharmaceutical composition for the treatment or prevention of atypical pneumonia.

20. A pharmaceutical composition comprising a virus according to any one of claim 1 to 5, a nucleic acid according to claim 6, a vector according to claim 7, a host cell according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9 an antigen according to claim 10, or an antibody according to claim 11.

21. A method for the treatment or prevention of an EMCR-CoV virus infection comprising providing an individual with a pharmaceutical composition according to claim 20.

22. A method for the treatment or prevention of atypical pneumonia comprising providing an individual with a pharmaceutical composition according to claim 20.

23. A viral replicase encoded by an RNA sequence comprising the indicated sequences, or homologues thereof as depicted in figure 1.

24. A viral spike protein comprising the indicated amino acid sequence as depicted figure 1, or a homologue thereof.

25 A viral nuclear capsid protein encoded by an RNA sequence comprising the indicated sequence as depicted in figure 1 or a homologue thereof.

26. A viral nsp 3 or envelope protein encoded by an RNA sequence comprising the indicated sequence as depicted in figure 1, or a homologue thereof.

27. A nucleic acid sequence which comprises one or more of the sequences coding for sepearte viral proteins as depicted in figure 1 or a nucleic acid sequence which can hybridise with any of these sequences under stringent conditions.

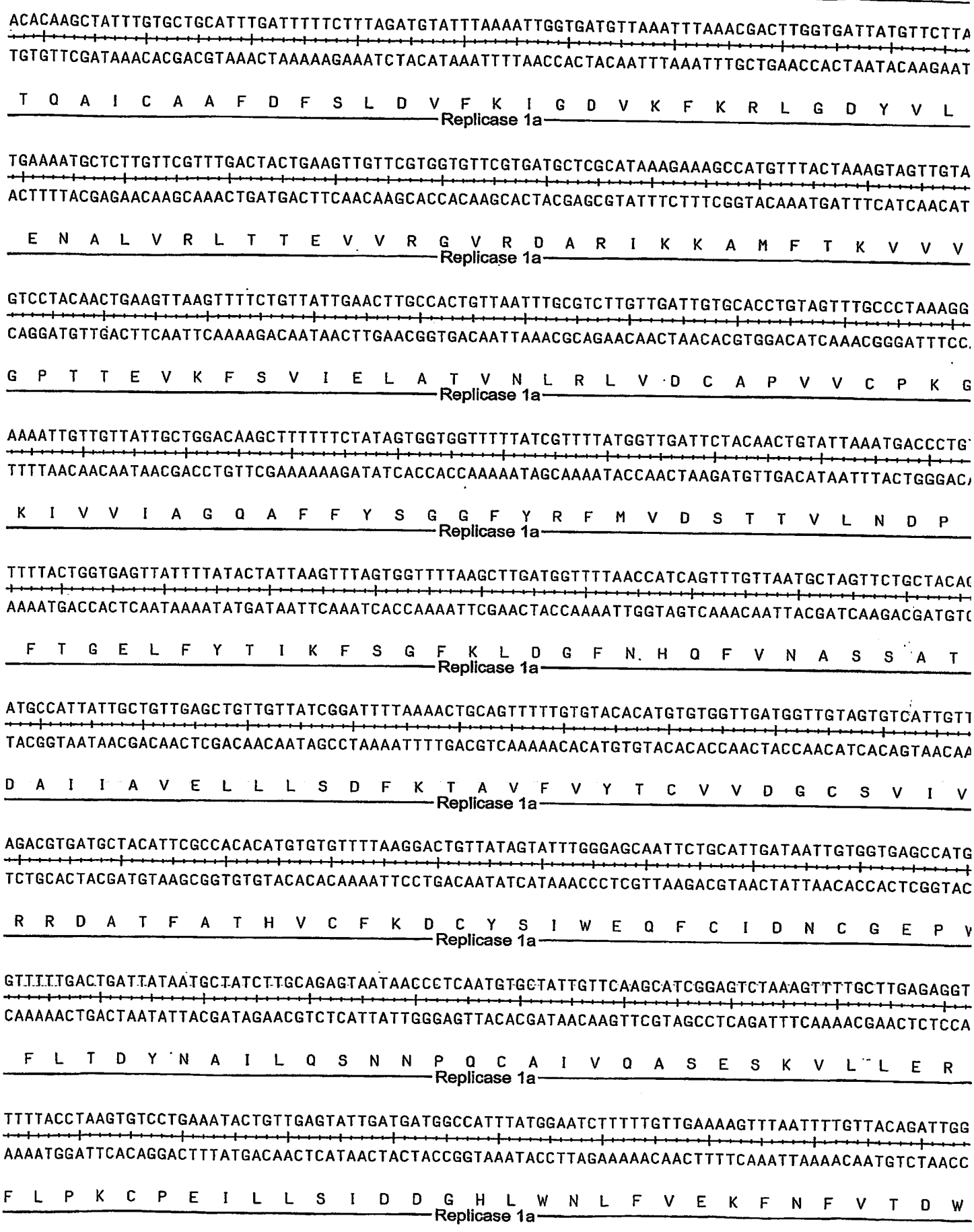
Abstract

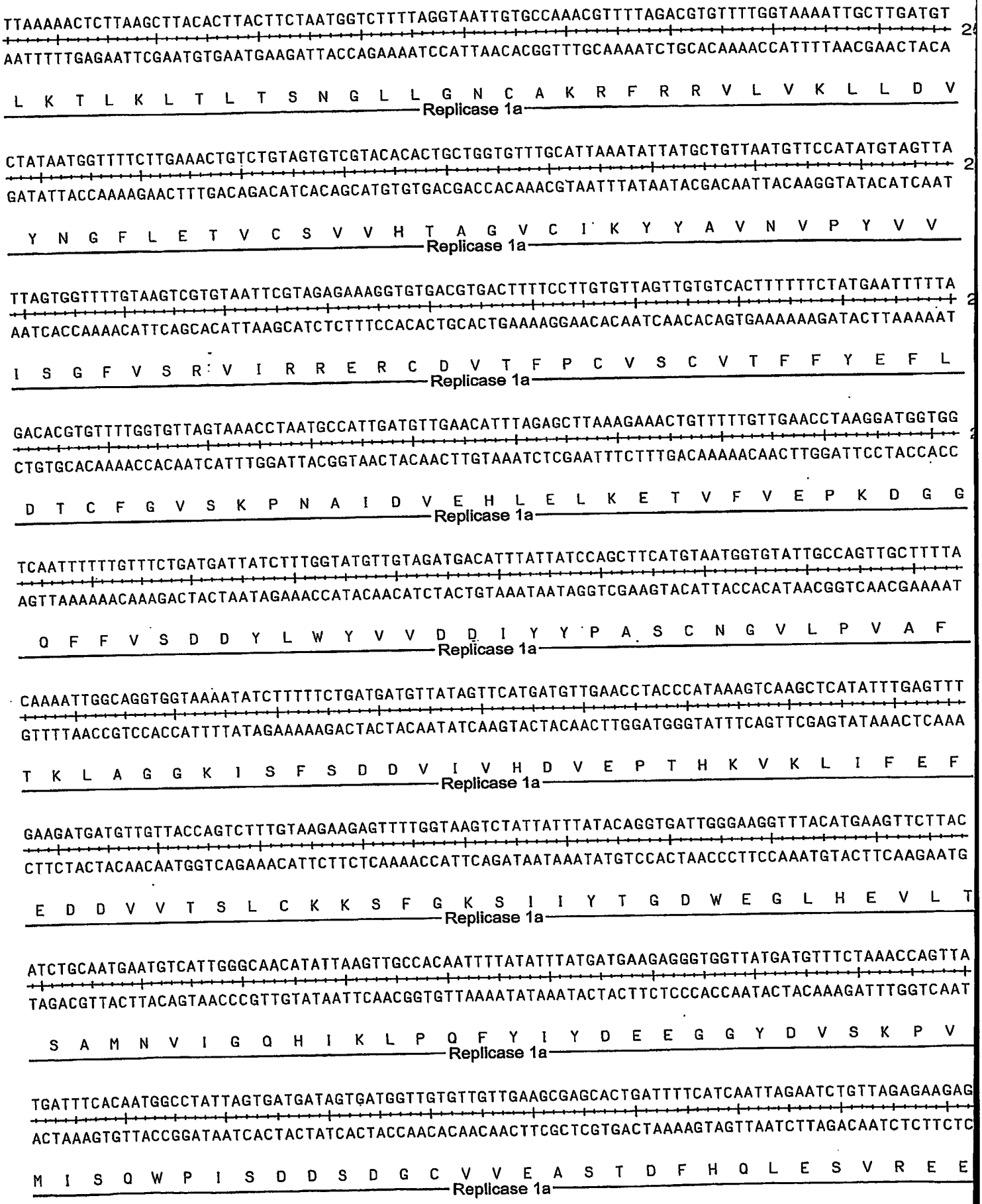
The invention relates to the field of virology. The invention provides a new isolated essentially mammalian positive-sense single stranded RNA virus (EMCR-CoV) within the group of coronaviuses and components thereof.

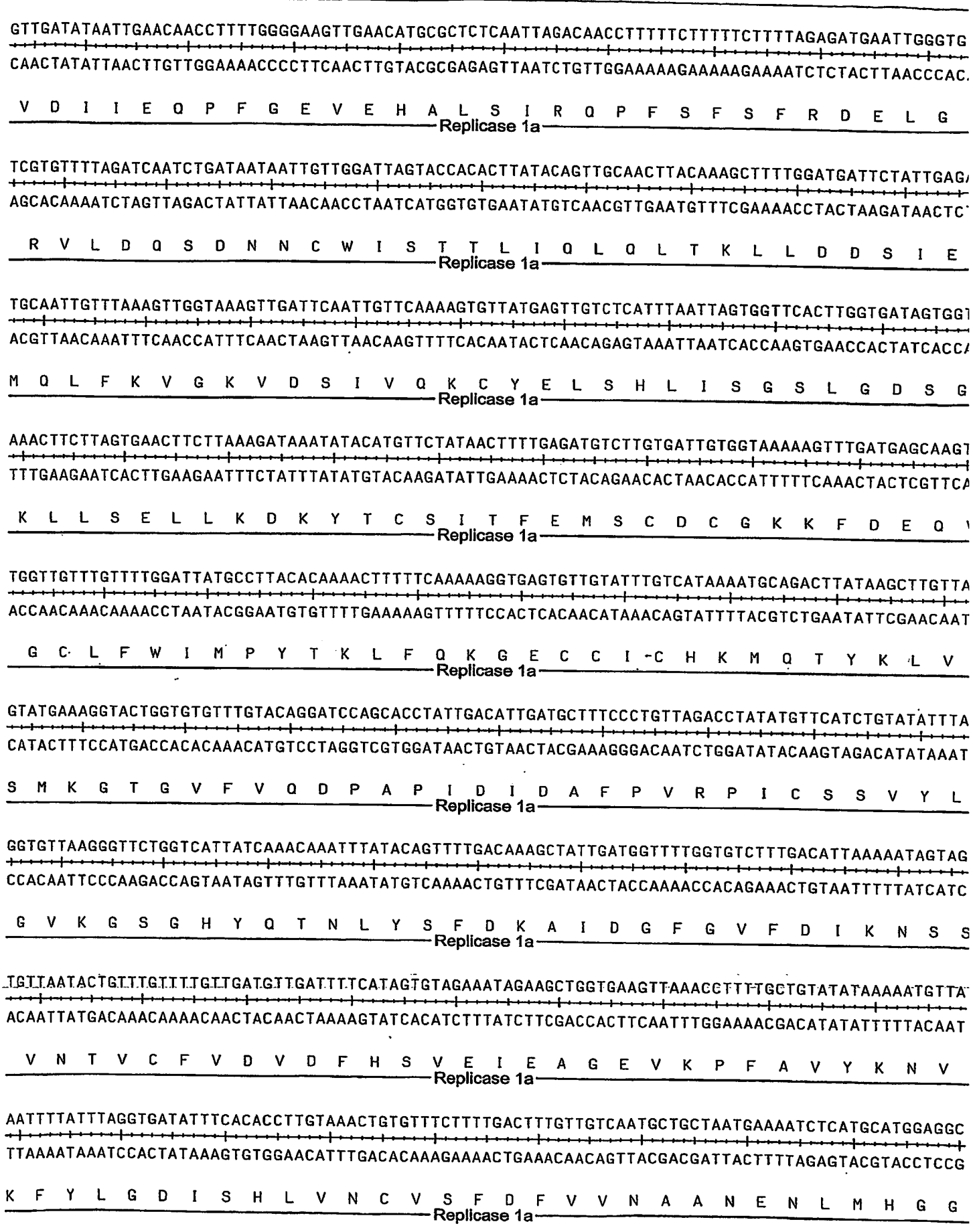
01 December 2003 11:52
EMCR-CoV.MPD (1 > 27532) Site and Sequence
Enzymes : All 212 enzymes (No Filter)
Settings : Circular, Certain Sites Only, Standard Genetic Code

```
AGATAGAGAATTTTCTTATTTAGACTTTGTGTCTACTCCTCTCAACTAAACGAAATTTTTCTAGTGCTGTCATTTGTTATGGCAGTCCTA
TCTATCTCTTAAAAGAATAAATCTGAAACACAGATGAGGAGAGTTGATTTGCTTTAAAAAGATCACGACAGTAAACAATACCGTCAGGAT

                    5'UTR

GTAATTGAAATTTCGTCAAGTTTGTAAACTGGTTAGGCAAGTGTTGTATTTTCTGTGTTTAAGCACTGGTGGTTCTGTCCACTAGTGCAC
CATTAACTTTAAAGCAGTTCAAACATTTGACCAATCCGTTCACAACATAAAAGACACAAATTCGTGACCACCAAGACAGGTGATCACGTG

                    5'UTR

ATTGATACTTAAGTGGTGTTCTGTCACTGCTTATTGTGGAAGCAACGTTCTGTCGTTGTGGAAACCAATAACTGCTAACCATGTTTTACA
TAACTATGAATTCACCACAAGACAGTGACGAATAACACCTTCGTTGCAAGACAGCAACACCTTTGGTTATTGACGATTGGTACAAAATGT

                    5'UTR                                   M F Y
                                                            Replicase 1a

CAAGTGACACTTGCTGTTGCAAGTGATTCGGAAATTTCAGGTTTTGGTTTTGCCATTCCTTCTGTAGCCGTTCGCGCTTATAGCGAAGCCG
GTTCACTGTGAACGACAACGTTCACTAAGCCTTTAAAGTCCAAAACCAAAACGGTAAGGAAGACATCGGCAAGCGCGAATATCGCTTCGGC

Q V T L A V A S D S E I S G F G F A I P S V A V R A Y S E A
                    Replicase 1a

TGCACAAGGTTTTCAGGCATGCCGCTTTGTTGCTTTTGGCTTACAGGATTGTGTAACCGGTATTAATGATGACGATTATGTCATTGCATTG
ACGTGTTCCAAAAGTCCGTACGGCGAAACAACGAAAACCGAATGTCCTAACACATTGGCCATAATTACTACTGCTAATACAGTAACGTAAC

A Q G F Q A C R F V A F G L Q D C V T G I N D D D Y V I A L
                    Replicase 1a

CTGGTACTAATCAGCTTTGTGCCAAAATTTTACTTTTTTCTGATAGACCTCTTAATTTGCGAGGTTGGCTCATTTTTTCTAACAGCAATTA
GACCATGATTAGTCGAAACACGGTTTTAAAATGAAAAAAGACTATCTGGAGAATTAAACGCTCCAACCGAGTAAAAAAGATTGTCGTTAAT

T G T N Q L C A K I L L F S D R P L N L R G W L I F S N S N Y
                    Replicase 1a

GTTCTTCAGGACTTTGATGTTGTTTTTGGCCATGGTGCAGGAAGTGTGGTTTTTGTGGATAAGTATATGTGTGGTTTTGATGGTAAACCTG
CAAGAAGTCCTGAAACTACAACAAAAACCGGTACCACGTCCTTCACACCAAAAACACCTATTCATATACACACCAAAACTACCATTTGGAC

V L Q D F D V V F G H G A G S V V F V D K Y M C G F D G K P
                    Replicase 1a

GTTACCTAAAAACATGTGGGAATTTAGAGATTACTTTAATGATAATACTGATAGTATTGTTATTGGTGGTGTCACTTATCAATTAGCATGGC
CAATGGATTTTTGTACACCCTTAAATCTCTAATGAAATTACTATTATGACTATCATAACAATAACCACCACAGTGAATAGTTAATCGTACCG

L P K N M W E F R D Y F N D N T D S I V I G G V T Y Q L A W
                    Replicase 1a

ATGTTATACGTAAAGACCTTTCTTATGAACAGCAAAATGTTTTAGCTATTGAGAGCATTCATTATCTTGGCACTACAGGTCATACTTTGAAG
TACAATATGCATTTCTGGAAAGAATACTTGTCGTTTTACAAAATCGATAACTCTCGTAAGTAATAGAACCGTGATGTCCAGTATGAAACTTC

D V I R K D L S Y E Q Q N V L A I E S I H Y L G T T G H T L K
                    Replicase 1a
```

```
TCTGGTTGCAAACTCATTAATGCCAAGCCGCCTAAATATTCTTCTAAGGTTGTTTTGAGTGGTGAATGGAATGCTGTGTATAAGGCGTTTGG
AGACCAACGTTTGAGTAATTACGGTTCGGCGGATTTATAAGAAGATTCCAACAAAACTCACCACTTACCTTACGACACATATTCCGCAAACC
 S  G  C  K  L  I  N  A  K  P  P  K  Y  S  S  K  V  V  L  S  G  E  W  N  A  V  Y  K  A  F  G
                                                Replicase 1a

TTCACCATTTATTACAAATGGTATATCATTGCTAGATATAATTGTTAAACCAGTTTTCTTTAATGCTTTTGTTAAATGCAATTGTGGTTCTG
AAGTGGTAAATAATGTTTACCATATAGTAACGATCTATATTAACAATTTGGTCAAAAGAAATTACGAAAACAATTTACGTTAACACCAAGAC
 S  P  F  I  T  N  G  I  S  L  L  D  I  I  V  K  P  V  F  F  N  A  F  V  K  C  N  C  G  S
                                                Replicase 1a

AGAATTGGAGTGTTGGTGCATGGGATGGTTATCTATCTTCTTGTTGTGGCACACCTGCTAAGAAACTTTGTGTTGTTCCTGGTAATGTTGTT
TCTTAACCTCACAACCACGTACCCTACCAATAGATAGAAGAACAACACCGTGTGGACGATTCTTTGAAACACAACAAGGACCATTACAACAA
 E  N  W  S  V  G  A  W  D  G  Y  L  S  S  C  C  G  T  P  A  K  K  L  C  V  V  P  G  N  V  V
                                                Replicase 1a

CCTGGTGATGTGATCATCACCTCAACTGATGCTGGTTGTGGTGTTAAATACTATGCTGGCTTAGTTGTTAAACATATTACTAACATTACTGG
GGACCACTACACTAGTAGTGGAGTTGACTACGACCAACACCACAATTTATGATACGACCGAATCAACAATTTGTATAATGATTGTAATGACC
 P  G  D  V  I  I  T  S  T  D  A  G  C  G  V  K  Y  Y  A  G  L  V  V  K  H  I  T  N  I  T  G
                                                Replicase 1a

TGTGTCTTTATGGCGTGTTACAGCTGTTCATTCTGATGGAATGTTTGTGGCAACATCTTCTTATGATGCACTTTTGCATAGAAATTCATTAG
ACACAGAAATACCGCACAATGTCGACAAGTAAGACTACCTTACAAACACCGTTGTAGAAGAATACTACGTGAAAACGTATCTTTAAGTAATC
 V  S  L  W  R  V  T  A  V  H  S  D  G  M  F  V  A  T  S  S  Y  D  A  L  L  H  R  N  S  L
                                                Replicase 1a

ACCCTTTTTGCTTTGATGTTAACACTTTACTTTCTAATCAATTACGTCTAGCTTTTCTTGGTGCTTCTGTTACAGAAGATGTTAAATTTGCT
TGGGAAAAACGAAACTACAATTGTGAAATGAAAGATTAGTTAATGCAGATCGAAAAGAACCACGAAGACAATGTCTTCTACAATTTAAACGA
 D  P  F  C  F  D  V  N  T  L  L  S  N  Q  L  R  L  A  F  L  G  A  S  V  T  E  D  V  K  F  A
                                                Replicase 1a

GCTAGCACTGGTGTTATTGACATTAGTGCTGGTATGTTTGGTCTTTACGATGACATATTGACAAACAATAAACCTTGGTTTGTACGCAAAGC
CGATCGTGACCACAATAACTGTAATCACGACCATACAAACCAGAAATGCTACTGTATAACTGTTTGTTATTTGGAACCAAACATGCGTTTCG
 A  S  T  G  V  I  D  I  S  A  G  M  F  G  L  Y  D  D  I  L  T  N  N  K  P  W  F  V  R  K  A
                                                Replicase 1a

TTCTGGGCTTTTTGATGCAATCTGGGATGCTTTTGTTGCCGCTATTAAGCTTGTGCCAACTACTACTGGTGGTTTGGTTAGGTTTGTTAAGT
AAGACCCGAAAAACTACGTTAGACCCTACGAAAACAACGGCGATAATTCGAACACGGTTGATGATGACCACCAAACCAATCCAAACAATTCA
 S  G  L  F  D  A  I  W  D  A  F  V  A  A  I  K  L  V  P  T  T  T  G  G  L  V  R  F  V  K
                                                Replicase 1a

CTATCGCTTCAACTGTTTTAACTGTTTCTAATGGTGTTATTATTATGTGTGCAGATGTTCCAGATGCTTTTCAACCAGTTTACCGCACATTT
GATAGCGAAGTTGACAAAATTGACAAAGATTACCACAATAATAATACACACGTCTACAAGGTCTACGAAAAGTTGGTCAAATGGCGTGTAAA
 S  I  A  S  T  V  L  T  V  S  N  G  V  I  I  M  C  A  D  V  P  D  A  F  Q  P  V  Y  R  T  F
                                                Replicase 1a
```


ACACAAGCTATTTGTGCTGCATTTGATTTTTCTTTAGATGTATTTAAAATTGGTGATGTTAAATTTAAACGACTTGGTGATTATGTTCTTA
TGTGTTCGATAAACACGACGTAAACTAAAAAGAAATCTACATAAATTTTAACCACTACAATTTAAATTTGCTGAACCACTAATACAAGAAT
T Q A I C A A F D F S L D V F K I G D V K F K R L G D Y V L
Replicase 1a
TGAAAATGCTCTTGTTCGTTTGACTACTGAAGTTGTTCGTGGTGTTCGTGATGCTCGCATAAAGAAAGCCATGTTTACTAAAGTAGTTGTA
ACTTTTACGAGAACAAGCAAACTGATGACTTCAACAAGCACCACAAGCACTACGAGCGTATTTCTTTCGGTACAAATGATTTCATCAACAT
E N A L V R L T T E V V R G V R D A R I K K A M F T K V V V
Replicase 1a
GTCCTACAACTGAAGTTAAGTTTTCTGTTATTGAACTTGCCACTGTTAATTTGCGTCTTGTTGATTGTGCACCTGTAGTTTGCCCTAAAGG
CAGGATGTTGACTTCAATTCAAAAGACAATAACTTGAACGGTGACAATTAAACGCAGAACAACTAACACGTGGACATCAAACGGGATTTCC
G P T T E V K F S V I E L A T V N L R L V D C A P V V C P K G
Replicase 1a
AAAATTGTTGTTATTGCTGGACAAGCTTTTTTCTATAGTGGTGGTTTTTATCGTTTTATGGTTGATTCTACAACTGTATTAAATGACCCTG
TTTTAACAACAATAACGACCTGTTCGAAAAAAGATATCACCACCAAAAATAGCAAAATACCAACTAAGATGTTGACATAATTTACTGGGAC
K I V V I A G Q A F F Y S G G F Y R F M V D S T T V L N D P
Replicase 1a
TTTTACTGGTGAGTTATTTTATACTATTAAGTTTAGTGGTTTTAAGCTTGATGGTTTTAACCATCAGTTTGTTAATGCTAGTTCTGCTACAC
AAAATGACCACTCAATAAAATATGATAATTCAAATCACCAAAATTCGAACTACCAAAATTGGTAGTCAAACAATTACGATCAAGACGATGTC
F T G E L F Y T I K F S G F K L D G F N H Q F V N A S S A T
Replicase 1a
ATGCCATTATTGCTGTTGAGCTGTTGTTATCGGATTTTAAAACTGCAGTTTTTGTGTACACATGTGTGGTTGATGGTTGTAGTGTCATTGTT
TACGGTAATAACGACAACTCGACAACAATAGCCTAAAATTTTGACGTCAAAAACACATGTGTACACACCAACTACCAACATCACAGTAACAA
D A I I A V E L L L S D F K T A V F V Y T C V V D G C S V I V
Replicase 1a
AGACGTGATGCTACATTCGCCACACATGTGTGTTTTAAGGACTGTTATAGTATTTGGGAGCAATTCTGCATTGATAATTGTGGTGAGCCATG
TCTGCACTACGATGTAAGCGGTGTGTACACACAAAATTCCTGACAATATCATAAACCCTCGTTAAGACGTAACTATTAACACCACTCGGTAC
R R D A T F A T H V C F K D C Y S I W E Q F C I D N C G E P V
Replicase 1a
GTTTTTGACTGATTATAATGCTATCTTGCAGAGTAATAACCCTCAATGTGGTATTGTTCAAGCATCGGAGTCTAAAGTTTTGCTTGAGAGGT
CAAAAACTGACTAATATTACGATAGAACGTCTCATTATTGGGAGTTACACGATAACAAGTTCGTAGCCTCAGATTTCAAAACGAACTCTCCA
F L T D Y N A I L Q S N N P Q C A I V Q A S E S K V L L E R
Replicase 1a
TTTTACCTAAGTGTCCTGAAATACTGTTGAGTATTGATGATGGCCATTTATGGAATCTTTTTGTTGAAAAGTTTAATTTTGTTACAGATTGG
AAAATGGATTCACAGGACTTTATGACAACTCATAACTACTACCGGTAAATACCTTAGAAAAACAACTTTTCAAATTAAAACAATGTCTAACC
F L P K C P E I L L S I D D G H L W N L F V E K F N F V T D W
Replicase 1a


TTAAAAACTCTTAAGCTTACACTTACTTCTAATGGTCTTTTAGGTAATTGTGCCAAACGTTTTAGACGTGTTTTGGTAAAATTGCTTGATGT
AATTTTTGAGAATTCGAATGTGAATGAAGATTACCAGAAAATCCATTAACACGGTTTGCAAAATCTGCACAAAACCATTTTAACGAACTACA
L K T L K L T L T S N G L L G N C A K R F R R V L V K L L D V
Replicase 1a
CTATAATGGTTTTCTTGAAACTGTCTGTAGTGTCGTACACACTGCTGGTGTTTGCATTAAATATTATGCTGTTAATGTTCCATATGTAGTTA
GATATTACCAAAAGAACTTTGACAGACATCACAGCATGTGTGACGACCACAAACGTAATTTATAATACGACAATTACAAGGTATACATCAAT
Y N G F L E T V C S V V H T A G V C I K Y Y A V N V P Y V V
Replicase 1a
TTAGTGGTTTTGTAAGTCGTGTAATTCGTAGAGAAAGGTGTGACGTGACTTTTCCTTGTGTTAGTTGTGTCACTTTTTTCTATGAATTTTTA
AATCACCAAAACATTCAGCACATTAAGCATCTCTTTCCACACTGCACTGAAAAGGAACACAATCAACACAGTGAAAAAAGATACTTAAAAAT
I S G F V S R V I R R E R C D V T F P C V S C V T F F Y E F L
Replicase 1a
GACACGTGTTTTGGTGTTAGTAAACCTAATGCCATTGATGTTGAACATTTAGAGCTTAAAGAAACTGTTTTTGTTGAACCTAAGGATGGTGG
CTGTGCACAAAACCACAATCATTTGGATTACGGTAACTACAACTTGTAAATCTCGAATTTCTTTGACAAAAACAACTTGGATTCCTACCACC
D T C F G V S K P N A I D V E H L E L K E T V F V E P K D G G
Replicase 1a
TCAATTTTTTGTTTCTGATGATTATCTTTGGTATGTTGTAGATGACATTTATTATCCAGCTTCATGTAATGGTGTATTGCCAGTTGCTTTTA
AGTTAAAAAACAAAGACTACTAATAGAAACCATACAACATCTACTGTAAATAATAGGTCGAAGTACATTACCACATAACGGTCAACGAAAAT
Q F F V S D D Y L W Y V V D D I Y Y P A S C N G V L P V A F
Replicase 1a
CAAAATTGGCAGGTGGTAAAATATCTTTTTCTGATGATGTTATAGTTCATGATGTTGAACCTACCCATAAAGTCAAGCTCATATTTGAGTTT
GTTTTAACCGTCCACCATTTTATAGAAAAAGACTACTACAATATCAAGTACTACAACTTGGATGGGTATTTCAGTTCGAGTATAAACTCAAA
T K L A G G K I S F S D D V I V H D V E P T H K V K L I F E F
Replicase 1a
GAAGATGATGTTGTTACCAGTCTTTGTAAGAAGAGTTTTGGTAAGTCTATTATTTATACAGGTGATTGGGAAGGTTTACATGAAGTTCTTAC
CTTCTACTACAACAATGGTCAGAAACATTCTTCTCAAAACCATTCAGATAATAAATATGTCCACTAACCCTTCCAAATGTACTTCAAGAATG
E D D V V T S L C K K S F G K S I I Y T G D W E G L H E V L T
Replicase 1a
ATCTGCAATGAATGTCATTGGGCAACATATTAAGTTGCCACAATTTTATATTTATGATGAAGAGGGTGGTTATGATGTTTCTAAACCAGTTA
TAGACGTTACTTACAGTAACCCGTTGTATAATTCAACGGTGTTAAAATATAAATACTACTTCTCCCACCAATACTACAAAGATTTGGTCAAT
S A M N V I G Q H I K L P Q F Y I Y D E E G G Y D V S K P V
Replicase 1a
TGATTTCACAATGGCCTATTAGTGATGATAGTGATGGTTGTGTTGTTGAAGCGAGCACTGATTTTCATCAATTAGAATCTGTTAGAGAAGAG
ACTAAAGTGTTACCGGATAATCACTACTATCACTACCAACACAACAACTTCGCTCGTGACTAAAAGTAGTTAATCTTAGACAATCTCTTCTC
M I S Q W P I S D D S D G C V V E A S T D F H Q L E S V R E E
Replicase 1a


GTTGATATAATTGAACAACCTTTTGGGGAAGTTGAACATGCGCTCTCAATTAGACAACCTTTTTCTTTTTCTTTTAGAGATGAATTGGGTG
CAACTATATTAACTTGTTGGAAAACCCCTTCAACTTGTACGCGAGAGTTAATCTGTTGGAAAAAGAAAAAGAAAATCTCTACTTAACCCAC
V D I I E Q P F G E V E H A L S I R Q P F S F S F R D E L G
Replicase 1a
TCGTGTTTTAGATCAATCTGATAATAATTGTTGGATTAGTACCACACTTATACAGTTGCAACTTACAAAGCTTTTGGATGATTCTATTGAG
AGCACAAAATCTAGTTAGACTATTATTAACAACCTAATCATGGTGTGAATATGTCAACGTTGAATGTTTCGAAAACCTACTAAGATAACTC
R V L D Q S D N N C W I S T T L I Q L Q L T K L L D D S I E
Replicase 1a
TGCAATTGTTTAAAGTTGGTAAAGTTGATTCAATTGTTCAAAAGTGTTATGAGTTGTCTCATTTAATTAGTGGTTCACTTGGTGATAGTGGT
ACGTTAACAAATTTCAACCATTTCAACTAAGTTAACAAGTTTTCACAATACTCAACAGAGTAAATTAATCACCAAGTGAACCACTATCACCA
M Q L F K V G K V D S I V Q K C Y E L S H L I S G S L G D S G
Replicase 1a
AAACTTCTTAGTGAACTTCTTAAAGATAAATATACATGTTCTATAACTTTTGAGATGTCTTGTGATTGTGGTAAAAAGTTTGATGAGCAAGT
TTTGAAGAATCACTTGAAGAATTTCTATTTATATGTACAAGATATTGAAAACTCTACAGAACACTAACACCATTTTTCAAACTACTCGTTCA
K L L S E L L K D K Y T C S I T F E M S C D C G K K F D E Q
Replicase 1a
TGGTTGTTTGTTTTGGATTATGCCTTACACAAAACTTTTTCAAAAAGGTGAGTGTTGTATTTGTCATAAAATGCAGACTTATAAGCTTGTTA
ACCAACAAACAAAACCTAATACGGAATGTGTTTTGAAAAAGTTTTTCCACTCACAACATAAACAGTATTTTACGTCTGAATATTCGAACAAT
G C L F W I M P Y T K L F Q K G E C C I C H K M Q T Y K L V
Replicase 1a
GTATGAAAGGTACTGGTGTGTTTGTACAGGATCCAGCACCTATTGACATTGATGCTTTCCCTGTTAGACCTATATGTTCATCTGTATATTTA
CATACTTTCCATGACCACACAAACATGTCCTAGGTCGTGGATAACTGTAACTACGAAAGGGACAATCTGGATATACAAGTAGACATATAAAT
S M K G T G V F V Q D P A P I D I D A F P V R P I C S S V Y L
Replicase 1a
GGTGTTAAGGGTTCTGGTCATTATCAAACAAATTTATACAGTTTTGACAAAGCTATTGATGGTTTTGGTGTCTTTGACATTAAAAATAGTAG
CCACAATTCCCAAGACCAGTAATAGTTTGTTTAAATATGTCAAAACTGTTTCGATAACTACCAAAACCACAGAAACTGTAATTTTTATCATC
G V K G S G H Y Q T N L Y S F D K A I D G F G V F D I K N S S
Replicase 1a
TGTTAATACTGTTTGTTTTGTTGATGTTGATTTTCATAGTGTAGAAATAGAAGCTGGTGAAGTTAAACCTTTTGCTGTATATAAAAATGTTA
ACAATTATGACAAACAAAACAACTACAACTAAAAGTATCACATCTTTATCTTCGACCACTTCAATTTGGAAAACGACATATATTTTTACAAT
V N T V C F V D V D F H S V E I E A G E V K P F A V Y K N V
Replicase 1a
AATTTTATTTAGGTGATATTTCACACCTTGTAAACTGTGTTTCTTTTGACTTTGTTGTCAATGCTGCTAATGAAAATCTCATGCATGGAGGC
TTAAAATAAATCCACTATAAAGTGTGGAACATTTGACACAAAGAAAACTGAAACAACAGTTACGACGATTACTTTTAGAGTACGTACCTCCG
K F Y L G D I S H L V N C V S F D F V V N A A N E N L M H G G
Replicase 1a

GGTGTCGCACGTGCTATTGATATTTTGACTGAAGGTCAACTTCAGTCATTATCTAAAGATTACATTAGTAGTAATGGTCCACTTAAGGTTGG 42
CCACAGCGTGCACGATAACTATAAAACTGACTTCCAGTTGAAGTCAGTAATAGATTTCTAATGTAATCATCATTACCAGGTGAATTCCAACC

G V A R A I D I L T E G Q L Q S L S K D Y I S S N G P L K V G
Replicase 1a

AGCAGGTGTTATGTTGGAGTGTGAAAAATTCAATGTATTTAATGTTGTTGGTCCGCGAACTGGTAAACATGAGCATTCATTACTTGTTGAAG 43
TCGTCCACAATACAACCTCACACTTTTTAAGTTACATAAATTACAACAACCAGGCGCTTGACCATTTGTACTCGTAAGTAATGAACAACTTC

A G V M L E C E K F N V F N V V G P R T G K H E H S L L V E
Replicase 1a

CTTATAATTCTATTTTATTTGAAAATGGTATTCCACTTATGCCTCTTCTTAGTTGTGGTATTTTTGGTGTAAGGATTGAAAATTCTCTTAAA 4
GAATATTAAGATAAAATAAACTTTTACCATAAGGTGAATACGGAGAAGAATCAACACCATAAAAACCACATTCCTAACTTTTAAGAGAATTT

A Y N S I L F E N G I P L M P L L S C G I F G V R I E N S L K
Replicase 1a

GCTTTGTTTAGTTGTGACATTAATAAACCATTGCAAGTTTTTGTTTATTCTTCAAATGAAGAACAAGCTGTTCTTAAGTTTTTAGATGGTTT 4
CGAAACAAATCAACACTGTAATTATTTGGTAACGTTCAAAAACAAATAAGAAGTTTACTTCTTGTTCGACAAGAATTCAAAAATCTACCAAA

A L F S C D I N K P L Q V F V Y S S N E E Q A V L K F L D G L
Replicase 1a

AGATTTAACACCAGTCATTGACGATGTTGATGTTGTTAAACCTTTTAGAGTTGAAGGTAATTTTTCATTCTTTGATTGTGGTGTCAATGCCT 4
TCTAAATTGTGGTCAGTAACTGCTACAACTACAACAATTTGGAAAATCTCAACTTCCATTAAAAAGTAAGAAACTAACACCACAGTTACGGA

D L T P V I D D V D V V K P F R V E G N F S F F D C G V N A
Replicase 1a

TGGATGGTGATATTTACTTATTATTTACTAACTCTATTTTAATGTTGGATAAACAAGGACAATTATTGGACACAAAACTTAATGGTATTTTG 4
ACCTACCACTATAAATGAATAATAAATGATTGAGATAAAATTACAACCTATTTGTTCCTGTTAATAACCTGTGTTTTGAATTACCATAAAAC

L D G D I Y L L F T N S I L M L D K Q G Q L L D T K L N G I L
Replicase 1a

CAACAGGCAGTTCTTGATTATCTTGCTACAGTTAAAACTGTACCAGCTGGTAATTTGGTTAAACTTGTTGTTGAGAGTTGTACCATTTATAT
GTTGTCCGTCAAGAACTAATAGAACGATGTCAATTTTGACATGGTCGACCATTAAACCAATTTGAACAACAACTCTCAACATGGTAAATATA

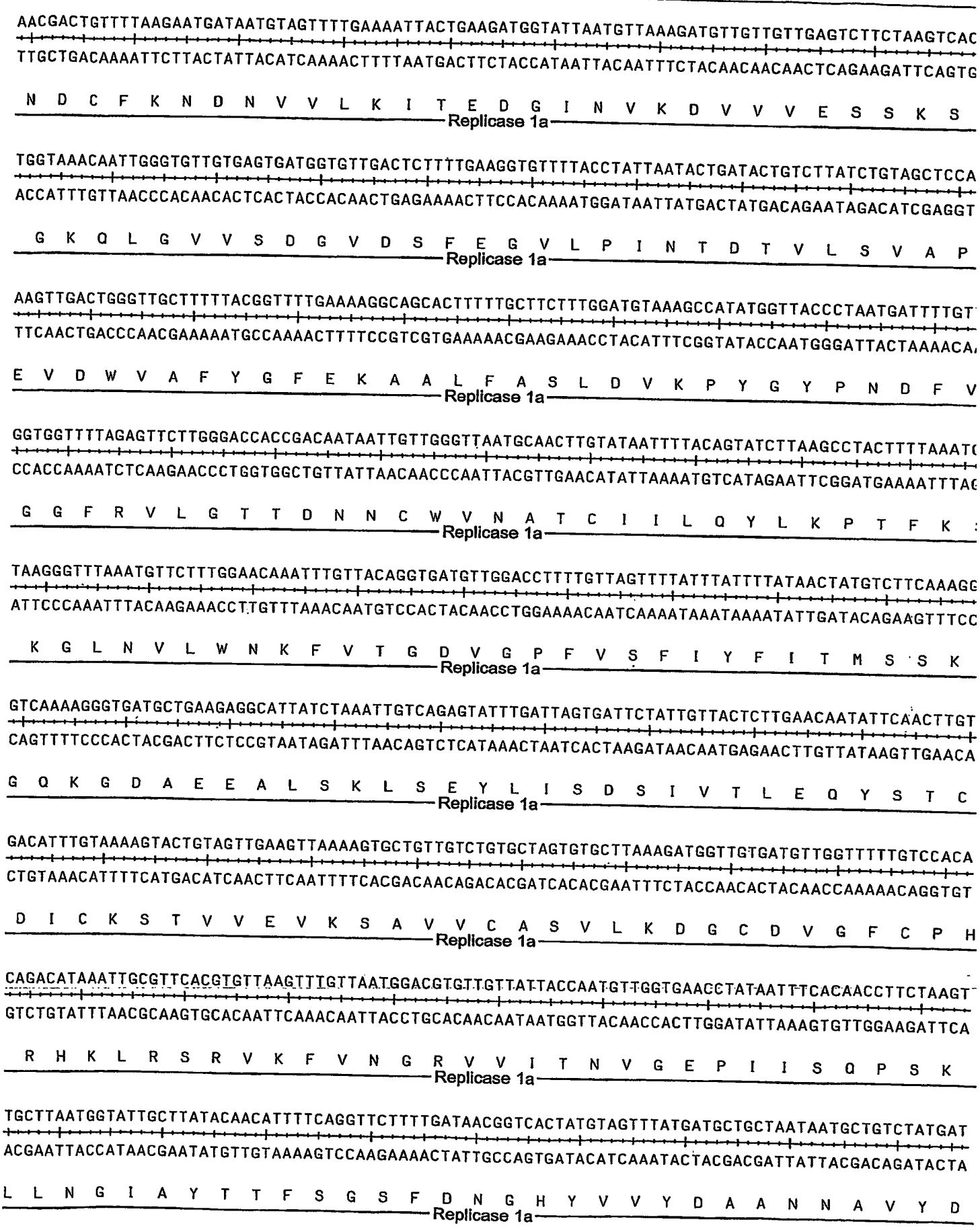
Q Q A V L D Y L A T V K T V P A G N L V K L V V E S C T I Y M
Replicase 1a

GTGTGTTGTACCATCGATAAATGATCTTTCTTTTGATAAAAATCTTGGTCGTTGTGTGCGTAAACTTAATAGATTGAAAACTTGTGTTATTG
CACACAACATGGTAGCTATTTACTAGAAAGAAAACTATTTTTAGAACCAGCAACACACGCATTTGAATTATCTAACTTTTGAACACAATAAC

C V V P S I N D L S F D K N L G R C V R K L N R L K T C V I
Replicase 1a

CCAATGTTCCTGCTATTGATGTTTTGAAAAAGCTTCTTTCAAGTTTGACTTTAACTGTTAAATTTGTTGTAGAGAGTAATGTTATGGATGTT
GGTTACAAGGACGATAACTACAAAACTTTTTCGAAGAAAGTTCAAACTGAAATTGACAATTTAAACAACATCTCTCATTACAATACCTACAA

A N V P A I D V L K K L L S S L T L T V K F V V E S N V M D V
Replicase 1a


AACGACTGTTTTAAGAATGATAATGTAGTTTTGAAAATTACTGAAGATGGTATTAATGTTAAAGATGTTGTTGTTGAGTCTTCTAAGTCAC
TTGCTGACAAAATTCTTACTATTACATCAAAACTTTTAATGACTTCTACCATAATTACAATTTCTACAACAACAACTCAGAAGATTCAGTG
N D C F K N D N V V L K I T E D G I N V K D V V V E S S K S
Replicase 1a
TGGTAAACAATTGGGTGTTGTGAGTGATGGTGTTGACTCTTTTGAAGGTGTTTTACCTATTAATACTGATACTGTCTTATCTGTAGCTCCA
ACCATTTGTTAACCCACAACACTCACTACCACAACTGAGAAAACTTCCACAAAATGGATAATTATGACTATGACAGAATAGACATCGAGGT
G K Q L G V V S D G V D S F E G V L P I N T D T V L S V A P
Replicase 1a
AAGTTGACTGGGTTGCTTTTTACGGTTTTGAAAAGGCAGCACTTTTTGCTTCTTTGGATGTAAAGCCATATGGTTACCCTAATGATTTTGT
TTCAACTGACCCAACGAAAAATGCCAAAACTTTTCCGTCGTGAAAAACGAAGAAACCTACATTTCGGTATACCAATGGGATTACTAAAACA
E V D W V A F Y G F E K A A L F A S L D V K P Y G Y P N D F V
Replicase 1a
GGTGGTTTTAGAGTTCTTGGGACCACCGACAATAATTGTTGGGTTAATGCAACTTGTATAATTTTACAGTATCTTAAGCCTACTTTTAAATC
CCACCAAAATCTCAAGAACCCTGGTGGCTGTTATTAACAACCCAATTACGTTGAACATATTAAAATGTCATAGAATTCGGATGAAAATTTAC
G G F R V L G T T D N N C W V N A T C I I L Q Y L K P T F K
Replicase 1a
TAAGGGTTTAAATGTTCTTTGGAACAAATTTGTTACAGGTGATGTTGGACCTTTTGTTAGTTTTATTTATTTTATAACTATGTCTTCAAAGG
ATTCCCAAATTTACAAGAAACCTTGTTTAAACAATGTCCACTACAACCTGGAAAACAATCAAAATAAATAAAATATTGATACAGAAGTTTCC
K G L N V L W N K F V T G D V G P F V S F I Y F I T M S S K
Replicase 1a
GTCAAAAGGGTGATGCTGAAGAGGCATTATCTAAATTGTCAGAGTATTTGATTAGTGATTCTATTGTTACTCTTGAACAATATTCAACTTGT
CAGTTTTCCCACTACGACTTCTCCGTAATAGATTTAACAGTCTCATAAACTAATCACTAAGATAACAATGAGAACTTGTTATAAGTTGAACA
G Q K G D A E E A L S K L S E Y L I S D S I V T L E Q Y S T C
Replicase 1a
GACATTTGTAAAAGTACTGTAGTTGAAGTTAAAAGTGCTGTTGTCTGTGCTAGTGTGCTTAAAGATGGTTGTGATGTTGGTTTTTGTCCACA
CTGTAAACATTTTCATGACATCAACTTCAATTTTCACGACAACAGACACGATCACACGAATTTCTACCAACACTACAACCAAAAACAGGTGT
D I C K S T V V E V K S A V V C A S V L K D G C D V G F C P H
Replicase 1a
CAGACATAAATTGCGTTCACGTGTTAAGTTTGTTAATGGACGTGTTGTTATTACCAATGTTGGTGAACCTATAATTTCACAACCTTCTAAGT
GTCTGTATTTAACGCAAGTGCACAATTCAAACAATTACCTGCACAACAATAATGGTTACAACCACTTGGATATTAAAGTGTTGGAAGATTCA
R H K L R S R V K F V N G R V V I T N V G E P I I S Q P S K
Replicase 1a
TGCTTAATGGTATTGCTTATACAACATTTTCAGGTTCTTTTGATAACGGTCACTATGTAGTTTATGATGCTGCTAATAATGCTGTCTATGAT
ACGAATTACCATAACGAATATGTTGTAAAAGTCCAAGAAAACTATTGCCAGTGATACATCAAATACTACGACGATTATTACGACAGATACTA
L L N G I A Y T T F S G S F D N G H Y V V Y D A A N N A V Y D
Replicase 1a

GGTGCTCGTTTATTTGCTTCAGATTTGTCTACTTTAGCTGTTACAGCTATTGTTGTAGTAGGTGGTTGTGTAACATCTAATGTTCCACCAAT 58
CCACGAGCAAATAAACGAAGTCTAAACAGATGAAATCGACAATGTCGATAACAACATCATCCACCAACACATTGTAGATTACAAGGTGGTTA

G A R L F A S D L S T L A V T A I V V V G G C V T S N V P P I
Replicase 1a

TGTTAGTGAGAAAATTTCTGTTATGGATAAACTTGATACTGGTGCACAAAAATTTTTCCAATTTGGTGATTTTGTTATGAATAACATTGTTC 5
ACAATCACTCTTTTAAAGACAATACCTATTTGAACTATGACCACGTGTTTTTAAAAAGGTTAAACCACTAAAACAATACTTATTGTAACAAG

V S E K I S V M D K L D T G A Q K F F Q F G D F V M N N I V
Replicase 1a

TGTTTTTAACTTGGTTGCTTAGTATGTTTAGTCTTTTACGTACTTCTATTATGAAGCATGATATTAAAGTTATTGCCAAGGCTCCTAAACGT 6
ACAAAAATTGAACCAACGAATCATACAAATCAGAAAATGCATGAAGATAATACTTCGTACTATAATTTCAATAACGGTTCCGAGGATTTGCA

L F L T W L L S M F S L L R T S I M K H D I K V I A K A P K R
Replicase 1a

ACAGGTGTTATTTTGACACGTAGTTTTAAGTATAACATTAGATCTGCTTTGTTTGTTGTAAAGCAGAAGTGGTGTGTTATTGTTACTTTGTT 6
TGTCCACAATAAAACTGTGCATCAAAATTCATATTGTAATCTAGACGAAACAAACAACATTTCGTCTTCACCACACAATAACAATGAAACAA

T G V I L T R S F K Y N I R S A L F V V K Q K W C V I V T L F
Replicase 1a

TAAGTTCTTATTGTTATTATATGCTATTTATGCACTTGTTTTTATGATTGTGCAATTTAGTCCTTTTAATAGTCTTTTATGTGGTGACATTG 6
ATTCAAGAATAACAATAATATACGATAAATACGTGAACAAAAATACTAACACGTTAAATCAGGAAAATTATCAGAAAATACACCACTGTAAC

K F L L L L Y A I Y A L V F M I V Q F S P F N S L L C G D I
Replicase 1a

TAAGTGGTTATGAAAAATCCACTTTTAATAAGGATATTTATTGTGGTAATTCTATGGTTTGTAAGATGTGTTTGTTTAGTTATCAAGAGTTT 6
ATTCACCAATACTTTTTAGGTGAAAATTATTCCTATAAATAACACCATTAAGATACCAAACATTCTACACAAACAAATCAATAGTTCTCAAA

V S G Y E K S T F N K D I Y C G N S M V C K M C L F S Y Q E F
Replicase 1a

AATGATTTGGATCATACTAGTCTTGTTTGGAAGCACATTCGTGATCCTATATTAATCAGTTTACAACCATTTGTTATACTTGTTATTTTGTT
TTACTAAACCTAGTATGATCAGAACAAACCTTCGTGTAAGCACTAGGATATAATTAGTCAAATGTTGGTAAACAATATGAACAATAAAACAA

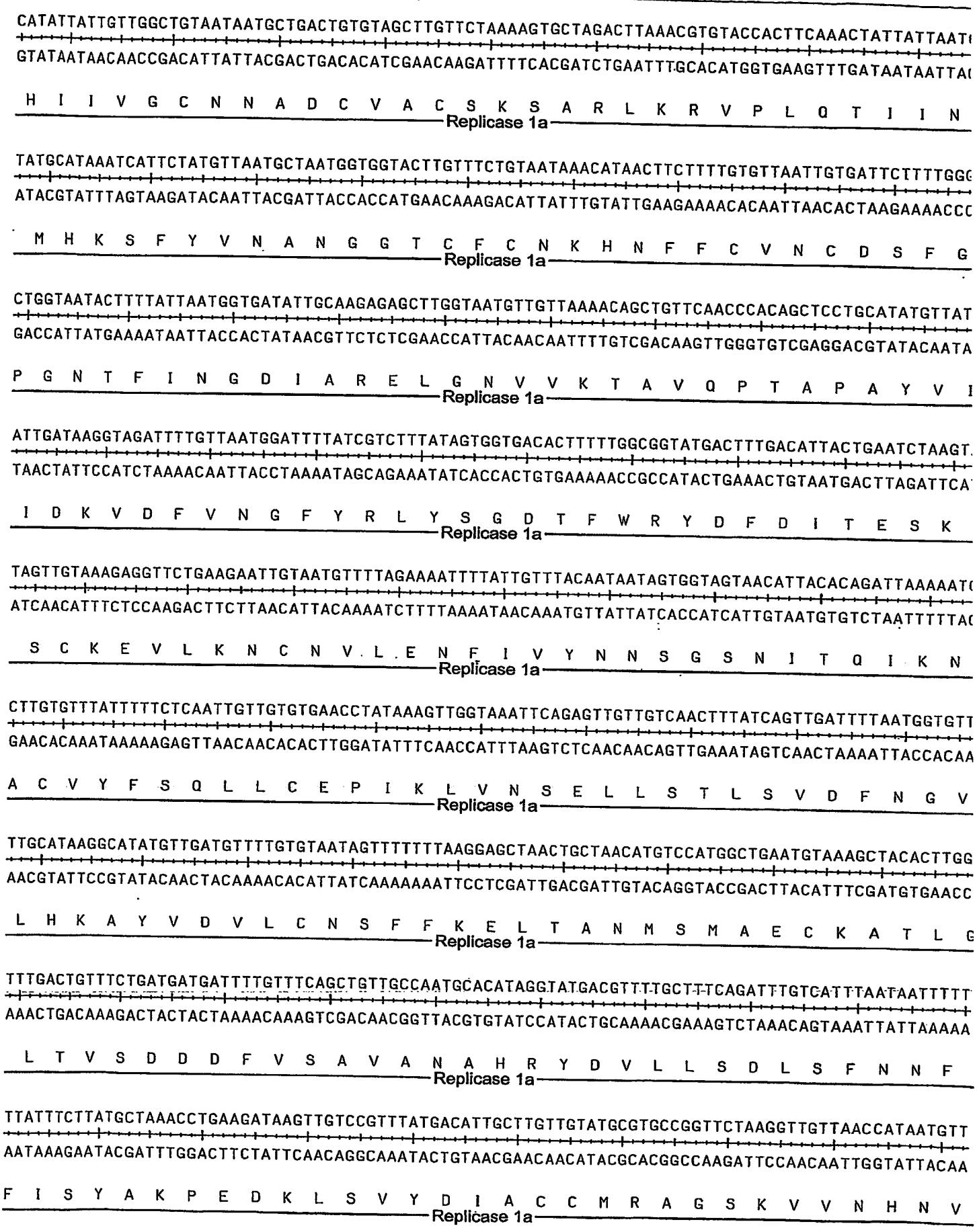
N D L D H T S L V W K H I R D P I L I S L Q P F V I L V I L L
Replicase 1a

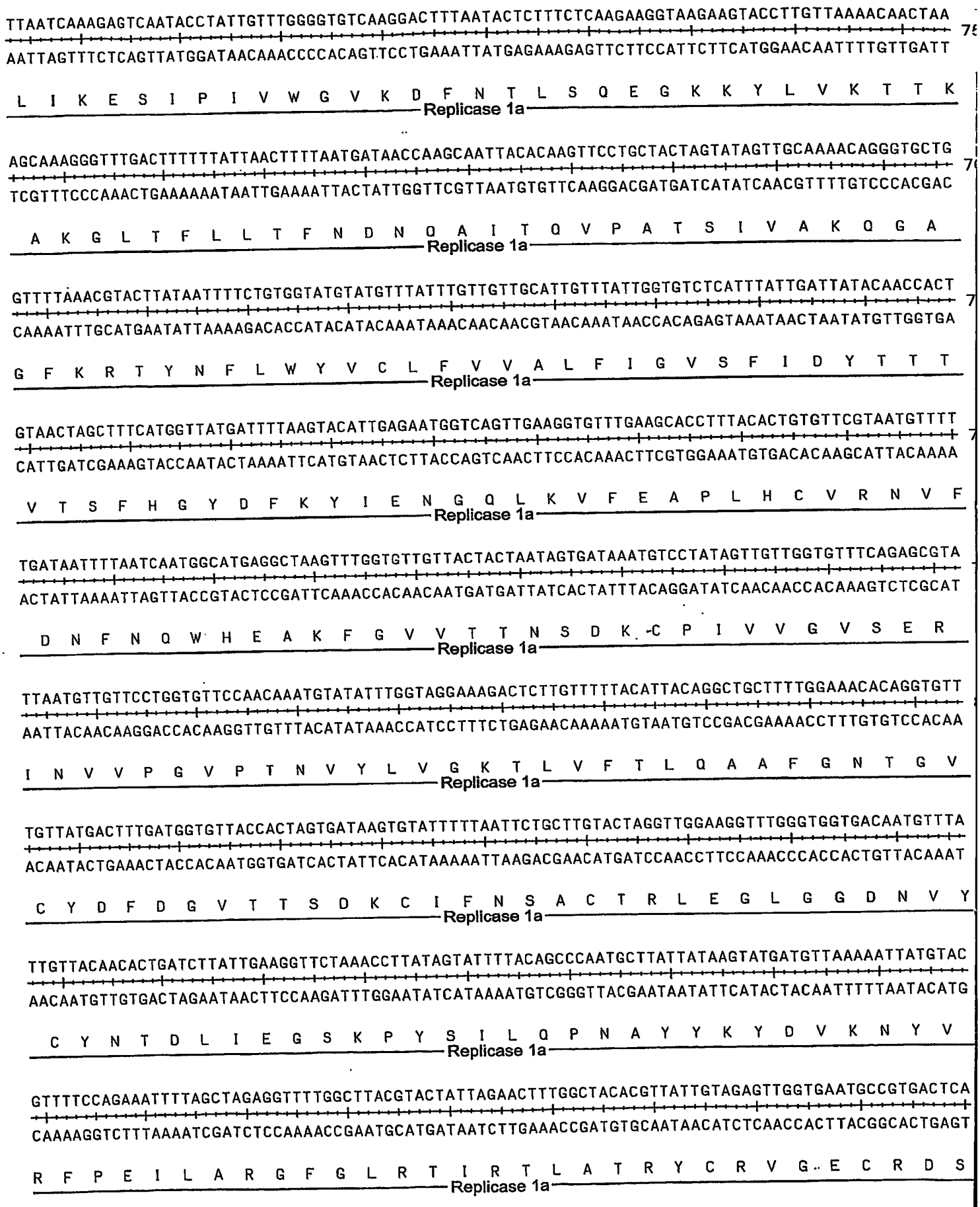
AATTTTTGGTAATATGTATTTGCGTTTTGGACTTTTATATTTTGTTGCACAATTTATTAGTACTTTTGGTTCTTTCTTAGGCTTTCATCAGA
TTAAAAACCATTATACATAAACGCAAAACCTGAAAATATAAAACAACGTGTTAAATAATCATGAAAACCAAGAAAGAATCCGAAAGTAGTCT

I F G N M Y L R F G L L Y F V A Q F I S T F G S F L G F H Q
Replicase 1a

AACAGTGGTTTTTACATTTTGTGCCGTTTGATGTTTTATGTAATGAGTTTTTAGCTACATTTATTGTCTGCAAAATTGTTTTATTTGTTAGA
TTGTCACCAAAAATGTAAAACACGGCAAACTACAAAATACATTACTCAAAAATCGATGTAAATAACAGACGTTTTAACAAAATAAACAATCT

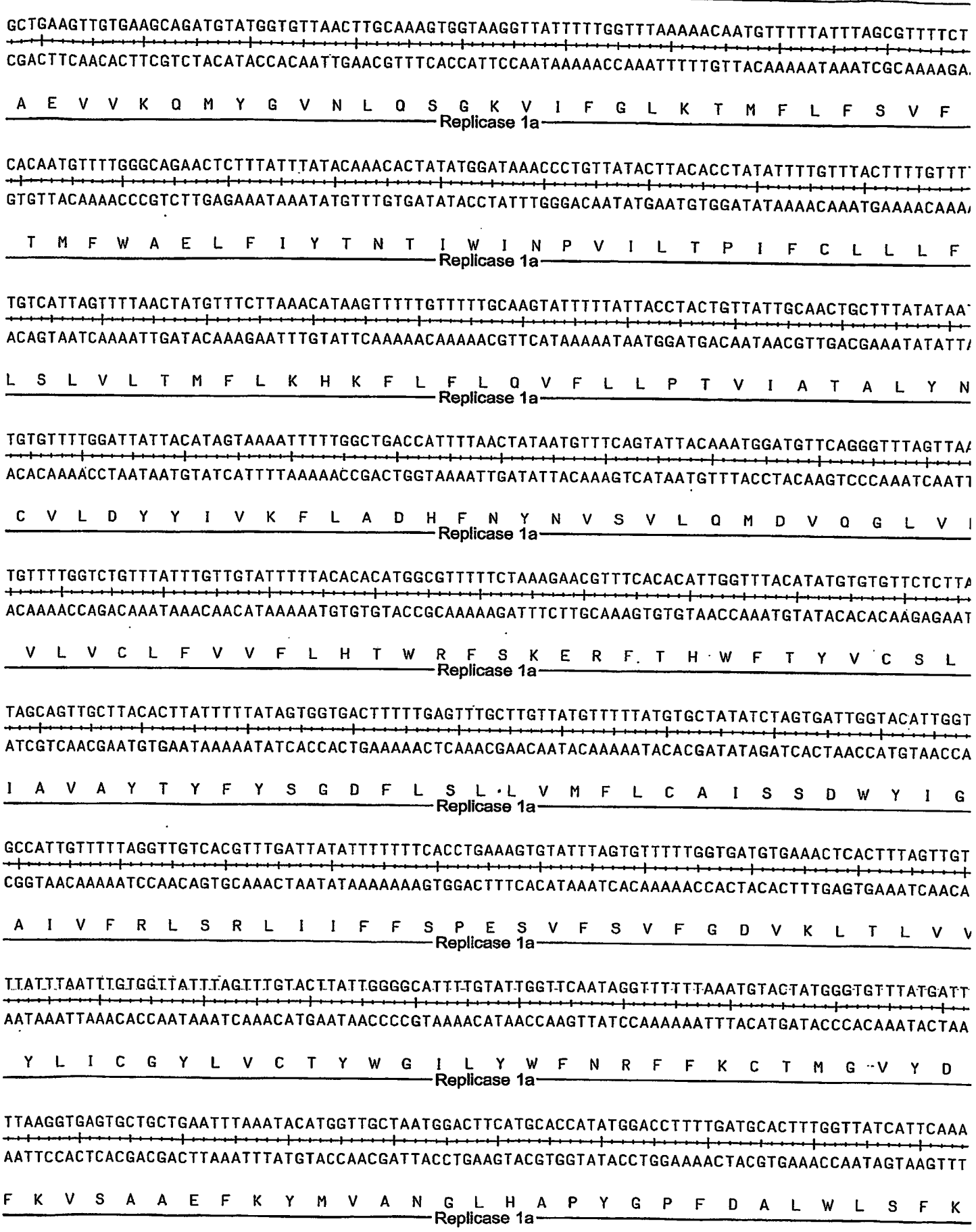
K Q W F L H F V P F D V L C N E F L A T F I V C K I V L F V R
Replicase 1a


CATATTATTGTTGGCTGTAATAATGCTGACTGTGTAGCTTGTTCTAAAAGTGCTAGACTTAAACGTGTACCACTTCAAACTATTATTAAT
GTATAATAACAACCGACATTATTACGACTGACACATCGAACAAGATTTTCACGATCTGAATTTGCACATGGTGAAGTTTGATAATAATTA
H I I V G C N N A D C V A C S K S A R L K R V P L Q T I I N
Replicase 1a
TATGCATAAATCATTCTATGTTAATGCTAATGGTGGTACTTGTTTCTGTAATAAACATAACTTCTTTTGTGTTAATTGTGATTCTTTTGG
ATACGTATTTAGTAAGATACAATTACGATTACCACCATGAACAAAGACATTATTTGTATTGAAGAAAACACAATTAACACTAAGAAAACC
M H K S F Y V N A N G G T C F C N K H N F F C V N C D S F G
Replicase 1a
CTGGTAATACTTTTATTAATGGTGATATTGCAAGAGAGCTTGGTAATGTTGTTAAAACAGCTGTTCAACCCACAGCTCCTGCATATGTTAT
GACCATTATGAAAATAATTACCACTATAACGTTCTCTCGAACCATTACAACAATTTTGTCGACAAGTTGGGTGTCGAGGACGTATACAATA
P G N T F I N G D I A R E L G N V V K T A V Q P T A P A Y V I
Replicase 1a
ATTGATAAGGTAGATTTTGTTAATGGATTTTATCGTCTTTATAGTGGTGACACTTTTTGGCGGTATGACTTTGACATTACTGAATCTAAGT
TAACTATTCCATCTAAAACAATTACCTAAAATAGCAGAAATATCACCACTGTGAAAAACCGCCATACTGAAACTGTAATGACTTAGATTCA
I D K V D F V N G F Y R L Y S G D T F W R Y D F D I T E S K
Replicase 1a
TAGTTGTAAAGAGGTTCTGAAGAATTGTAATGTTTTAGAAAATTTTATTGTTTACAATAATAGTGGTAGTAACATTACACAGATTAAAAAT
ATCAACATTTCTCCAAGACTTCTTAACATTACAAAATCTTTTAAAATAACAAATGTTATTATCACCATCATTGTAATGTGTCTAATTTTTA
S C K E V L K N C N V L E N F I V Y N N S G S N I T Q I K N
Replicase 1a
CTTGTGTTTATTTTTCTCAATTGTTGTGTGAACCTATAAAGTTGGTAAATTCAGAGTTGTTGTCAACTTTATCAGTTGATTTTAATGGTGTT
GAACACAAATAAAAAGAGTTAACAACACACTTGGATATTTCAACCATTTAAGTCTCAACAACAGTTGAAATAGTCAACTAAAATTACCACAA
A C V Y F S Q L L C E P I K L V N S E L L S T L S V D F N G V
Replicase 1a
TTGCATAAGGCATATGTTGATGTTTTGTGTAATAGTTTTTTTAAGGAGCTAACTGCTAACATGTCCATGGCTGAATGTAAAGCTACACTTGG
AACGTATTCCGTATACAACTACAAAACACATTATCAAAAAAATTCCTCGATTGACGATTGTACAGGTACCGACTTACATTTCGATGTGAACC
L H K A Y V D V L C N S F F K E L T A N M S M A E C K A T L G
Replicase 1a
TTTGACTGTTTCTGATGATGATTTTGTTTCAGCTGTTGCCAATGCACATAGGTATGACGTTTTGCTTTCAGATTTGTCATTTAATAATTTTT
AAACTGACAAAGACTACTACTAAAACAAAGTCGACAACGGTTACGTGTATCCATACTGCAAAACGAAAGTCTAAACAGTAAATTATTAAAAA
L T V S D D D F V S A V A N A H R Y D V L L S D L S F N N F
Replicase 1a
TTATTTCTTATGCTAAACCTGAAGATAAGTTGTCCGTTTATGACATTGCTTGTTGTATGCGTGCCGGTTCTAAGGTTGTTAACCATAATGTT
AATAAAGAATACGATTTGGACTTCTATTCAACAGGCAAATACTGTAACGAACAACATACGCACGGCCAAGATTCCAACAATTGGTATTACAA
F I S Y A K P E D K L S V Y D I A C C M R A G S K V V N H N V
Replicase 1a


TTAATCAAAGAGTCAATACCTATTGTTTGGGGTGTCAAGGACTTTAATACTCTTTCTCAAGAAGGTAAGAAGTACCTTGTTAAAACAACTAA
AATTAGTTTCTCAGTTATGGATAACAAACCCCACAGTTCCTGAAATTATGAGAAAGAGTTCTTCCATTCTTCATGGAACAATTTTGTTGATT
L I K E S I P I V W G V K D F N T L S Q E G K K Y L V K T T K
Replicase 1a
AGCAAAGGGTTTGACTTTTTTATTAACTTTTAATGATAACCAAGCAATTACACAAGTTCCTGCTACTAGTATAGTTGCAAAACAGGGTGCTG
TCGTTTCCCAAACTGAAAAAATAATTGAAAATTACTATTGGTTCGTTAATGTGTTCAAGGACGATGATCATATCAACGTTTTGTCCCACGAC
A K G L T F L L T F N D N Q A I T Q V P A T S I V A K Q G A
Replicase 1a
GTTTTAAACGTACTTATAATTTTCTGTGGTATGTATGTTTATTTGTTGTTGCATTGTTTATTGGTGTCTCATTTATTGATTATACAACCACT
CAAAATTTGCATGAATATTAAAAGACACCATACATACAAATAAACAACAACGTAACAAATAACCACAGAGTAAATAACTAATATGTTGGTGA
G F K R T Y N F L W Y V C L F V V A L F I G V S F I D Y T T T
Replicase 1a
GTAACTAGCTTTCATGGTTATGATTTTAAGTACATTGAGAATGGTCAGTTGAAGGTGTTTGAAGCACCTTTACACTGTGTTCGTAATGTTTT
CATTGATCGAAAGTACCAATACTAAAATTCATGTAACTCTTACCAGTCAACTTCCACAAACTTCGTGGAAATGTGACACAAGCATTACAAAA
V T S F H G Y D F K Y I E N G Q L K V F E A P L H C V R N V F
Replicase 1a
TGATAATTTTAATCAATGGCATGAGGCTAAGTTTGGTGTTGTTACTACTAATAGTGATAAATGTCCTATAGTTGTTGGTGTTTCAGAGCGTA
ACTATTAAAATTAGTTACCGTACTCCGATTCAAACCACAACAATGATGATTATCACTATTTACAGGATATCAACAACCACAAAGTCTCGCAT
D N F N Q W H E A K F G V V T T N S D K C P I V V G V S E R
Replicase 1a
TTAATGTTGTTCCTGGTGTTCCAACAAATGTATATTTGGTAGGAAAGACTCTTGTTTTTACATTACAGGCTGCTTTTGGAAACACAGGTGTT
AATTACAACAAGGACCACAAGGTTGTTTACATATAAACCATCCTTTCTGAGAACAAAAATGTAATGTCCGACGAAAACCTTTGTGTCCACAA
I N V V P G V P T N V Y L V G K T L V F T L Q A A F G N T G V
Replicase 1a
TGTTATGACTTTGATGGTGTTACCACTAGTGATAAGTGTATTTTTAATTCTGCTTGTACTAGGTTGGAAGGTTTGGGTGGTGACAATGTTTA
ACAATACTGAAACTACCACAATGGTGATCACTATTCACATAAAAATTAAGACGAACATGATCCAACCTTCCAAACCCACCACTGTTACAAAT
C Y D F D G V T T S D K C I F N S A C T R L E G L G G D N V Y
Replicase 1a
TTGTTACAACACTGATCTTATTGAAGGTTCTAAACCTTATAGTATTTTACAGCCCAATGCTTATTATAAGTATGATGTTAAAAATTATGTAC
AACAATGTTGTGACTAGAATAACTTCCAAGATTTGGAATATCATAAAATGTCGGGTTACGAATAATATTCATACTACAATTTTTAATACATG
C Y N T D L I E G S K P Y S I L Q P N A Y Y K Y D V K N Y V
Replicase 1a
GTTTTCCAGAAATTTTAGCTAGAGGTTTTGGCTTACGTACTATTAGAACTTTGGCTACACGTTATTGTAGAGTTGGTGAATGCCGTGACTCA
CAAAAGGTCTTTAAAATCGATCTCCAAAACCGAATGCATGATAATCTTGAAACCGATGTGCAATAACATCTCAACCACTTACGGCACTGAGT
R F P E I L A R G F G L R T I R T L A T R Y C R V G E C R D S
Replicase 1a

```
CATAAAGGTGTTTGTTTTGGTTTTGATAAATGGTATGTTAATGATGGACGTGTTGATGACGGTTACATTTGTGGTGATGGTCTTATAGACC
GTATTTCCACAAACAAAACCAAAACTATTTACCATACAATTACTACCTGCACAACTACTGCCAATGTAAACACCACTACCAGAATATCTGG

 H  K  G  V  C  F  G  F  D  K  W  Y  V  N  D  G  R  V  D  D  G  Y  I  C  G  D  G  L  I  D
                                  Replicase 1a

TCTTGTTAATGTACTCTCAATCTTTAGTTCATCTTTTAGCGTTGTGGCTATGTCTGGACATATGTTGTTTAATTTTCTTTTTGCAGCATTT
AGAACAATTACATGAGAGTTAGAAATCAAGTAGAAAATCGCAACACCGATACAGACCTGTATACAACAAATTAAAAGAAAAACGTCGTAAA

  L  V  N  V  L  S  I  F  S  S  S  F  S  V  V  A  M  S  G  H  M  L  F  N  F  L  F  A  A  F
                                  Replicase 1a

TTACATTTTTGTGCTTTTTAGTTACTAAATTTAAACGTGTTTTTGGTGATCTTTCTTATGGTGTTTTTACTGTTGTTTGTGCAACTTTGAT
AATGTAAAAACACGAAAAATCAATGATTTAAATTTGCACAAAAACCACTAGAAAGAATACCACAAAAATGACAACAAACACGTTGAAACTA

 I  T  F  L  C  F  L  V  T  K  F  K  R  V  F  G  D  L  S  Y  G  V  F  T  V  V  C  A  T  L  I
                                  Replicase 1a

AATAACATTTCTTATGTTGTTACTCAAAATTTATTTTTTATGTTGCTTTATGCTATTTTGTATTTTGTTTTTTACTAGGACAGTGCGTTATG
TTATTGTAAAGAATACAACAATGAGTTTTAAATAAAAAATACAACGAAATACGATAAAACATAAAACAAAAATGATCCTGTCACGCAATAC

 N  N  I  S  Y  V  V  T  Q  N  L  F  F  M  L  L  Y  A  I  L  Y  F  V  F  T  R  T  V  R  Y
                                  Replicase 1a

TTGGATTTGGCATATTGCATACATTGTTGCATACTTCTTGTTAATACCATGGTGGCTTCTCACATGGTTTAGTTTTGCTGCATTTTTAGAG
AACCTAAACCGTATAACGTATGTAACAACGTATGAAGAACAATTATGGTACCACCGAAGAGTGTACCAAATCAAAACGACGTAAAAATCTC

  W  I  W  H  I  A  Y  I  V  A  Y  F  L  L  I  P  W  W  L  L  T  W  F  S  F  A  A  F  L  E
                                  Replicase 1a

TTTTACCTAATGTTTTTAAGTTAAAAATCTCTACTCAATTGTTTGAAGGTGATAAGTTTATAGGTACTTTTGAGAGTGCTGCTGCAGGTACA
AAAATGGATTACAAAAATTCAATTTTTAGAGATGAGTTAACAAACTTCCACTATTCAAATATCCATGAAAACTCTCACGACGACGTCCATGT

 L  L  P  N  V  F  K  L  K  I  S  T  Q  L  F  E  G  D  K  F  I  G  T  F  E  S  A  A  A  G  T
                                  Replicase 1a

TTTGTTCTTGACATGCGTTCTTATGAAAGGCTGATAAATACTATTTCACCTGAGAAACTTAAGAATTATGCTGCAAGTTATAATAAATATAA
AAACAAGAACTGTACGCAAGAATACTTTCCGACTATTTATGATAAAGTGGACTCTTTGAATTCTTAATACGACGTTCAATATTATTTATATT

 F  V  L  D  M  R  S  Y  E  R  L  I  N  T  I  S  P  E  K  L  K  N  Y  A  A  S  Y  N  K  Y  I
                                  Replicase 1a

ATATTATAGTGGTAGTGCTAGTGAGGCTGATTATCGTTGTGCTTGTTATGCTCATTTAGCCAAGGCTATGTTAGATTACGCAAAAGATCATA
TATAATATCACCATCACGATCACTCCGACTAATAGCAACACGAACAATACGAGTAAATCGGTTCCGATACAATCTAATGCGTTTTCTAGTAT

  Y  Y  S  G  S  A  S  E  A  D  Y  R  C  A  C  Y  A  H  L  A  K  A  M  L  D  Y  A  K  D  H
                                  Replicase 1a

ATGACATGTTATATTCTCCACCTACCATTAGCTACAATTCCACCTTACAATCTGGTCTTAAGAAGATGGCACAACCATCTGGTTGTGTTGAG
TACTGTACAATATAAGAGGTGGATGGTAATCGATGTTAAGGTGGAATGTTAGACCAGAATTCTTCTACCGTGTTGGTAGACCAACACAACTC

 N  D  M  L  Y  S  P  P  T  I  S  Y  N  S  T  L  Q  S  G  L  K  K  M  A  Q  P  S  G  C  V  E
                                  Replicase 1a
```

```
AGATGTGTGGTTCGCGTCTGTTATGGTAGTACTGTGCTTAATGGAGTTTGGTTAGGTGACACTGTTACTTGTCCTAGACATGTCATAGCACC
TCTACACACCAAGCGCAGACAATACCATCATGACACGAATTACCTCAAACCAATCCACTGTGACAATGAACAGGATCTGTACAGTATCGTGG

 R  C  V  V  R  V  C  Y  G  S  T  V  L  N  G  V  W  L  G  D  T  V  T  C  P  R  H  V  I  A  P
                                        Replicase 1a

ATCAACCACTGTTCTTATTGATTATGATCATGCATATAGTACTATGCGTTTGCATAATTTTTCAGTGTCTCATAATGGTGTCTTCTTGGGAG
TAGTTGGTGACAAGAATAACTAATACTAGTACGTATATCATGATACGCAAACGTATTAAAAAGTCACAGAGTATTACCACAGAAGAACCCTC

 S  T  T  V  L  I  D  Y  D  H  A  Y  S  T  M  R  L  H  N  F  S  V  S  H  N  G  V  F  L  G
                                        Replicase 1a

TTGTTGGTGTTACAATGCATGGTTCTGTGTTGCGTATTAAGGTTTCACAATCTAATGTACATACACCTAAACATGTTTTTAAAACGTTGAAA
AACAACCACAATGTTACGTACCAAGACACAACGCATAATTCCAAAGTGTTAGATTACATGTATGTGGATTTGTACAAAAATTTTGCAACTTT

 V  V  G  V  T  M  H  G  S  V  L  R  I  K  V  S  Q  S  N  V  H  T  P  K  H  V  F  K  T  L  K
                                        Replicase 1a

CCTGGTGCTTCTTTTAATATTTTAGCATGTTATGAAGGTATTGCATCTGGTGTTTTTGGTGTTAATTTACGTACAAACTTTACTATTAAAGG
GGACCACGAAGAAAATTATAAAATCGTACAATACTTCCATAACGTAGACCACAAAAACCACAATTAAATGCATGTTTGAAATGATAATTTCC

 P  G  A  S  F  N  I  L  A  C  Y  E  G  I  A  S  G  V  F  G  V  N  L  R  T  N  F  T  I  K  G
                                        Replicase 1a

TTCTTTTATAAATGGAGCTTGTGGTTCTCCTGGTTATAATGTTAGAAATGATGGTACTGTTGAGTTTTGTTATTTACACCAAATTGAGTTAG
AAGAAAATATTTACCTCGAACACCAAGAGGACCAATATTACAATCTTTACTACCATGACAACTCAAAACAATAAATGTGGTTTAACTCAATC

 S  F  I  N  G  A  C  G  S  P  G  Y  N  V  R  N  D  G  T  V  E  F  C  Y  L  H  Q  I  E  L
                                        Replicase 1a

GTAGTGGTGCTCATGTTGGTTCTGATTTTACTGGTAGTGTTTATGGTAATTTTGATGACCAACCTAGTTTGCAAGTTGAGAGTGCCAACCTT
CATCACCACGAGTACAACCAAGACTAAAATGACCATCACAAATACCATTAAAACTACTGGTTGGATCAAACGTTCAACTCTCACGGTTGGAA

 G  S  G  A  H  V  G  S  D  F  T  G  S  V  Y  G  N  F  D  D  Q  P  S  L  Q  V  E  S  A  N  L
                                        Replicase 1a

ATGCTATCAGATAATGTTGTTGCCTTTTTGTATGCTGCTTTGTTGAATGGTTGTAGGTGGTGGTTGCGTTCAACTAGAGTTAATGTTGATGG
TACGATAGTCTATTACAACAACGGAAAAACATACGACGAAACAACTTACCAACATCCACCACCAACGCAAGTTGATCTCAATTACAACTACC

 M  L  S  D  N  V  V  A  F  L  Y  A  A  L  L  N  G  C  R  W  W  L  R  S  T  R  V  N  V  D  G
                                        Replicase 1a

TTTTAATGAATGGGCTATGGCTAATGGTTATACAATTGTTTCTAGTGTTGAGTGCTATTCTATTTTGGCAGCAAAAACTGGTGTTAGTGTTG
AAAATTACTTACCCGATACCGATTACCAATATGTTAACAAAGATCACAACTCACGATAAGATAAAACCGTCGTTTTTGACCACAATCACAAC

 F  N  E  W  A  M  A  N  G  Y  T  I  V  S  S  V  E  C  Y  S  I  L  A  A  K  T  G  V  S  V
                                        Replicase 1a

AACAATTGTTAGCTTCCATTCAACATCTTCATGAAGGTTTTGGTGGTAAAAACATACTTGGTTATTCTAGTTTATGTGATGAGTTCACACTA
TTGTTAACAATCGAAGGTAAGTTGTAGAAGTACTTCCAAAACCACCATTTTTGTATGAACCAATAAGATCAAATACACTACTCAAGTGTGAT

 E  Q  L  L  A  S  I  Q  H  L  H  E  G  F  G  G  K  N  I  L  G  Y  S  S  L  C  D  E  F  T  L
                                        Replicase 1a
```


GCTGAAGTTGTGAAGCAGATGTATGGTGTTAACTTGCAAAGTGGTAAGGTTATTTTTGGTTTAAAAACAATGTTTTTATTTAGCGTTTTCT
CGACTTCAACACTTCGTCTACATACCACAATTGAACGTTTCACCATTCCAATAAAAACCAAATTTTTGTTACAAAAATAAATCGCAAAAGA
A E V V K Q M Y G V N L Q S G K V I F G L K T M F L F S V F
Replicase 1a
CACAATGTTTTGGGCAGAACTCTTTATTTATACAAACACTATATGGATAAACCCTGTTATACTTACACCTATATTTTGTTTACTTTTGTTT
GTGTTACAAAACCCGTCTTGAGAAATAAATATGTTTGTGATATACCTATTTGGGACAATATGAATGTGGATATAAAACAAATGAAAACAAA
T M F W A E L F I Y T N T I W I N P V I L T P I F C L L L F
Replicase 1a
TGTCATTAGTTTTAACTATGTTTCTTAAACATAAGTTTTTGTTTTTGCAAGTATTTTTATTACCTACTGTTATTGCAACTGCTTTATATAA
ACAGTAATCAAAATTGATACAAAGAATTTGTATTCAAAAACAAAAACGTTCATAAAAATAATGGATGACAATAACGTTGACGAAATATATT
L S L V L T M F L K H K F L F L Q V F L L P T V I A T A L Y N
Replicase 1a
TGTGTTTTGGATTATTACATAGTAAAATTTTTGGCTGACCATTTTAACTATAATGTTTCAGTATTACAAATGGATGTTCAGGGTTTAGTTA
ACACAAAACCTAATAATGTATCATTTTAAAAACCGACTGGTAAAATTGATATTACAAAGTCATAATGTTTACCTACAAGTCCCAAATCAATT
C V L D Y Y I V K F L A D H F N Y N V S V L Q M D V Q G L V
Replicase 1a
TGTTTTGGTCTGTTTATTTGTTGTATTTTTACACACATGGCGTTTTTCTAAAGAACGTTTCACACATTGGTTTACATATGTGTGTTCTCTTA
ACAAAACCAGACAAATAAACAACATAAAAATGTGTGTACCGCAAAAAGATTTCTTGCAAAGTGTGTAACCAAATGTATACACACAAGAGAAT
V L V C L F V V F L H T W R F S K E R F T H W F T Y V C S L
Replicase 1a
TAGCAGTTGCTTACACTTATTTTTATAGTGGTGACTTTTTGAGTTTGCTTGTTATGTTTTTATGTGCTATATCTAGTGATTGGTACATTGGT
ATCGTCAACGAATGTGAATAAAAATATCACCACTGAAAAACTCAAACGAACAATACAAAAATACACGATATAGATCACTAACCATGTAACCA
I A V A Y T Y F Y S G D F L S L L V M F L C A I S S D W Y I G
Replicase 1a
GCCATTGTTTTTAGGTTGTCACGTTTGATTATATTTTTTTCACCTGAAAGTGTATTTAGTGTTTTTGGTGATGTGAAACTCACTTTAGTTGT
CGGTAACAAAAATCCAACAGTGCAAACTAATATAAAAAAAGTGGACTTTCACATAAATCACAAAAACCACTACACTTTGAGTGAAATCAACA
A I V F R L S R L I I F F S P E S V F S V F G D V K L T L V V
Replicase 1a
TTATTTAATTTGTGGTTATTTAGTTTGTACTTATTGGGGCATTTTGTATTGGTTCAATAGGTTTTTTAAATGTACTATGGGTGTTTATGATT
AATAAATTAAACACCAATAAATCAAACATGAATAACCCCGTAAAACATAACCAAGTTATCCAAAAAATTTACATGATACCCACAAATACTAA
Y L I C G Y L V C T Y W G I L Y W F N R F F K C T M G V Y D
Replicase 1a
TTAAGGTGAGTGCTGCTGAATTTAAATACATGGTTGCTAATGGACTTCATGCACCATATGGACCTTTTGATGCACTTTGGTTATCATTCAAA
AATTCCACTCACGACGACTTAAATTTATGTACCAACGATTACCTGAAGTACGTGGTATACCTGGAAAACTACGTGAAACCAATAGTAAGTTT
F K V S A A E F K Y M V A N G L H A P Y G P F D A L W L S F K
Replicase 1a

TTACTTGGTATTGGTGGTGACCGTTGTATAAAAATTTCAACTGTCCAATCCAAACTGACTGATTTGAAGTGTACTAATGTTGTGTTATTGGG 1
AATGAACCATAACCACCACTGGCAACATATTTTTAAAGTTGACAGGTTAGGTTTGACTGACTAAACTTCACATGATTACAACACAATAACCC

L L G I G G D R C I K I S T V Q S K L T D L K C T N V V L L G
Replicase 1a

TTGTTTGTCTAGTATGAACATTGCAGCTAATTCTAGTGAATGGGCTTATTGTGTTGATTTACACAATAAGATTAATCTTTGTGATGACCCAG 1
AACAAACAGATCATACTTGTAACGTCGATTAAGATCACTTACCCGAATAACACAACTAAATGTGTTATTCTAATTAGAAACACTACTGGGTC

C L S S M N I A A N S S E W A Y C V D L H N K I N L C D D P
Replicase 1a

AAAAAGCTCAAGGTATGTTGTTAGCACTCCTTGCGTTCTTTCTAAGTAAACATAGTGATTTTGGTCTTGATGGCCTTATTGATTCTTATTTT
TTTTTCGAGTTCCATACAACAATCGTGAGGAACGCAAGAAAGATTCATTTGTATCACTAAAACCAGAACTACCGGAATAACTAAGAATAAAA

E K A Q G M L L A L L A F F L S K H S D F G L D G L I D S Y F
Replicase 1a

GATAATAGTAGCACCCTGCAGAGTGTTGCTTCATCATTTGTTAGTATGCCATCATATATTGCTTATGAAAATGCTAGACAAGCTTATGAGGA
CTATTATCATCGTGGGACGTCTCACAACGAAGTAGTAAACAATCATACGGTAGTATATAACGAATACTTTTACGATCTGTTCGAATACTCCT

D N S S T L Q S V A S S F V S M P S Y I A Y E N A R Q A Y E D
Replicase 1a

TGCTATTGCTAATGGATCTTCTTCTCAACTTATTAAACAATTGAAGCGTGCCATGAATATCGCAAAGTCTGAATTTGATCATGAGATATCTG
ACGATAACGATTACCTAGAAGAAGAGTTGAATAATTTGTTAACTTCGCACGGTACTTATAGCGTTTCAGACTTAAACTAGTACTCTATAGAC

A I A N G S S S Q L I K Q L K R A M N I A K S E F D H E I S
Replicase 1a

TTCAGAAGAAAATTAATAGAATGGCTGAACAAGCTGCTACTCAGATGTATAAAGAAGCACGCTCTGTTAATAGAAAATCTAAAGTTATTAGT
AAGTCTTCTTTTAATTATCTTACCGACTTGTTCGACGATGAGTCTACATATTTCTTCGTGCGAGACAATTATCTTTTAGATTTCAATAATCA

V Q K K I N R M A E Q A A T Q M Y K E A R S V N R K S K V I S
Replicase 1a

GCTATGCACTCTTTACTTTTTGGAATGTTAAGACGTTTGGATATGTCTAGTGTTGAAACTGTTTTGAATTTAGCACGTGATGGTGTTGTGCC
CGATACGTGAGAAATGAAAAACCTTACAATTCTGCAAACCTATACAGATCACAACTTTGACAAAACTTAAATCGTGCACTACCACAACACGG

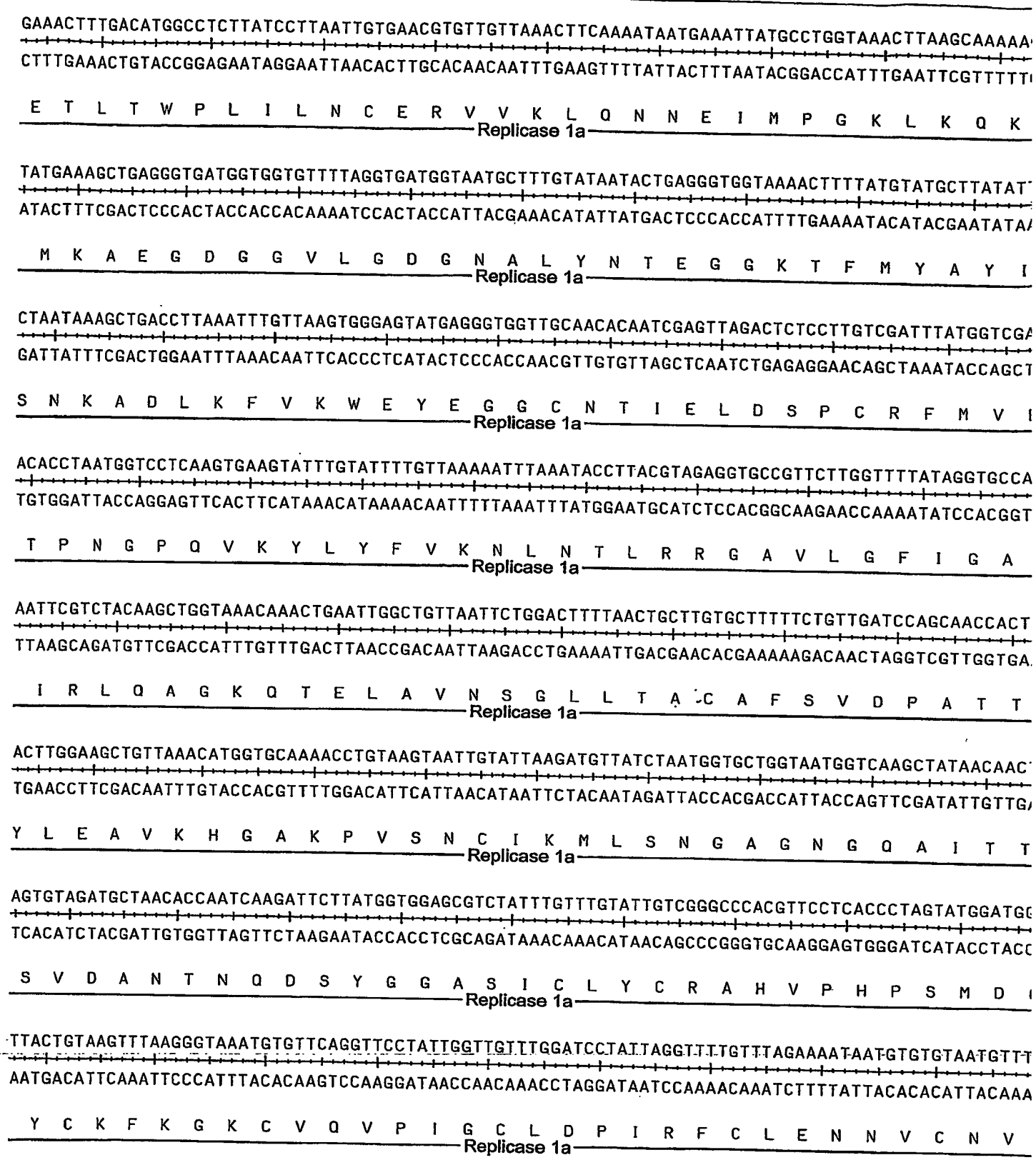
A M H S L L F G M L R R L D M S S V E T V L N L A R D G V V P
Replicase 1a

ATTGTCAGTTATACCTGCAACTTCAGCTTCCAAACTAACTATTGTTAGTCCAGATCTTGAATCTTATTCTAAGATTGTTTGTGATGGTTCTG
TAACAGTCAATATGGACGTTGAAGTCGAAGGTTTGATTGATAACAATCAGGTCTAGAACTTAGAATAAGATTCTAACAAACACTACCAAGAC

L S V I P A T S A S K L T I V S P D L E S Y S K I V C D G S
Replicase 1a

TTCATTATGCTGGAGTTGTTTGGACACTTAATGATGTTAAAGACAATGATGGTAGACCTGTTCATGTTAAAGAGATTACAAGGGAGAATGTT
AAGTAATACGACCTCAACAAACCTGTGAATTACTACAATTTCTGTTACTACCATCTGGACAAGTACAATTTCTCTAATGTTCCCTCTTACAA

V H Y A G V V W T L N D V K D N D G R P V H V K E I T R E N V
Replicase 1a


GAAACTTTGACATGGCCTCTTATCCTTAATTGTGAACGTGTTGTTAAACTTCAAAATAATGAAATTATGCCTGGTAAACTTAAGCAAAAA
CTTTGAAACTGTACCGGAGAATAGGAATTAACACTTGCACAACAATTTGAAGTTTTATTACTTTAATACGGACCATTTGAATTCGTTTTT
E T L T W P L I L N C E R V V K L Q N N E I M P G K L K Q K
Replicase 1a
TATGAAAGCTGAGGGTGATGGTGGTGTTTTAGGTGATGGTAATGCTTTGTATAATACTGAGGGTGGTAAAACTTTTATGTATGCTTATAT
ATACTTTCGACTCCCACTACCACCACAAAATCCACTACCATTACGAAACATATTATGACTCCCACCATTTTGAAAATACATACGAATATA
M K A E G D G G V L G D G N A L Y N T E G G K T F M Y A Y I
Replicase 1a
CTAATAAAGCTGACCTTAAATTTGTTAAGTGGGAGTATGAGGGTGGTTGCAACACAATCGAGTTAGACTCTCCTTGTCGATTTATGGTCGA
GATTATTTCGACTGGAATTTAAACAATTCACCCTCATACTCCCACCAACGTTGTGTTAGCTCAATCTGAGAGGAACAGCTAAATACCAGCT
S N K A D L K F V K W E Y E G G C N T I E L D S P C R F M V
Replicase 1a
ACACCTAATGGTCCTCAAGTGAAGTATTTGTATTTTGTTAAAAATTTAAATACCTTACGTAGAGGTGCCGTTCTTGGTTTTATAGGTGCCA
TGTGGATTACCAGGAGTTCACTTCATAAACATAAAACAATTTTTAAATTTATGGAATGCATCTCCACGGCAAGAACCAAAATATCCACGGT
T P N G P Q V K Y L Y F V K N L N T L R R G A V L G F I G A
Replicase 1a
AATTCGTCTACAAGCTGGTAAACAAACTGAATTGGCTGTTAATTCTGGACTTTTAACTGCTTGTGCTTTTTCTGTTGATCCAGCAACCACT
TTAAGCAGATGTTCGACCATTTGTTTGACTTAACCGACAATTAAGACCTGAAAATTGACGAACACGAAAAAGACAACTAGGTCGTTGGTGA
I R L Q A G K Q T E L A V N S G L L T A C A F S V D P A T T
Replicase 1a
ACTTGGAAGCTGTTAAACATGGTGCAAAACCTGTAAGTAATTGTATTAAGATGTTATCTAATGGTGCTGGTAATGGTCAAGCTATAACAAC
TGAACCTTCGACAATTTGTACCACGTTTTGGACATTCATTAACATAATTCTACAATAGATTACCACGACCATTACCAGTTCGATATTGTTG
Y L E A V K H G A K P V S N C I K M L S N G A G N G Q A I T T
Replicase 1a
AGTGTAGATGCTAACACCAATCAAGATTCTTATGGTGGAGCGTCTATTTGTTTGTATTGTCGGGCCCACGTTCCTCACCCTAGTATGGATGG
TCACATCTACGATTGTGGTTAGTTCTAAGAATACCACCTCGCAGATAAACAAACATAACAGCCCGGGTGCAAGGAGTGGGATCATACCTACC
S V D A N T N Q D S Y G G A S I C L Y C R A H V P H P S M D
Replicase 1a
TTACTGTAAGTTTAAGGGTAAATGTGTTCAGGTTCCTATTGGTTGTTTGGATCCTATTAGGTTTTGTTTAGAAAATAATGTGTGTAATGTTT
AATGACATTCAAATTCCCATTTACACAAGTCCAAGGATAACCAACAAACCTAGGATAATCCAAAACAAATCTTTTATTACACACATTACAAA
Y C K F K G K C V Q V P I G C L D P I R F C L E N N V C N V
Replicase 1a

GTGGTTGTTGGTTGGGACACGGGTGTGCTTGTGATCGTACAACCATTCAAAGTGTTGACATTTCTTATTTAAACGAGCAAGGGGTTCTAGTG 12
CACCAACAACCAACCCTGTGCCCACACGAACACTAGCATGTTGGTAAGTTTCACAACTGTAAAGAATAAATTTGCTCGTTCCCCAAGATCAC

C G C W L G H G C A C D R T T I Q S V D I S Y L N E Q G V L V
Replicase 1a

R A R G S S
Replicase 1b

CAGCTCGACTAGAACCCTGTAATGGCACGGACATCGATAAGTGTGTTCGTGCTTTTGACATTTATAATAAAAATGTTTCATTCTTGGGTAAG 1:
GTCGAGCTGATCTTGGGACATTACCGTGCCTGTAGCTATTCACACAAGCACGAAAACTGTAAATATTATTTTTACAAAGTAAGAACCCATTC

Q L D
Replicase 1a

A A R L E P C N G T D I D K C V R A F D I Y N K N V S F L G K
Replicase 1b

TGTTTGAAGATGAACTGTGTTCGTTTTAAAAATGCTGATCTTAAGGATGGTTATTTTGTTATAAAGAGGTGTACTAAGTCGGTTATGGAACA 1
ACAAACTTCTACTTGACACAAGCAAAATTTTTACGACTAGAATTCCTACCAATAAAACAATATTTCTCCACATGATTCAGCCAATACCTTGT

C L K M N C V R F K N A D L K D G Y F V I K R C T K S V M E H
Replicase 1b

CGAGCAATCCATGTATAACCTACTTAACTTTTCTGGTGCTTTGGCTGAGCATGATTTCTTTACTTGGAAAGATGGCAGAGTCATTTATGGTA
GCTCGTTAGGTACATATTGGATGAATTGAAAAGACCACGAAACCGACTCGTACTAAAGAAATGAACCTTTCTACCGTCTCAGTAAATACCAT

E Q S M Y N L L N F S G A L A E H D F F T W K D G R V I Y G
Replicase 1b

ATGTTAGTAGACATAATCTTACTAAATATACTATGATGGACTTGGTTTATGCTATGCGTAACTTTGATGAACAAAATTGTGATGTTCTAAAA
TACAATCATCTGTATTAGAATGATTTATATGATACTACCTGAACCAAATACGATACGCATTGAAACTACTTGTTTTAACACTACAAGATTTT

N V S R H N L T K Y T M M D L V Y A M R N F D E Q N C D V L K
Replicase 1b

GAAGTATTAGTTTTAACTGGTTGTTGTGACAATTCTTATTTTGATAGTAAGGGTTGGTATGACCCAGTTGAAAATGAAGATATACATAGAGT
CTTCATAATCAAAATTGACCAACAACACTGTTAAGAATAAAACTATCATTCCCAACCATACTGGGTCAACTTTTACTTCTATATGTATCTCA

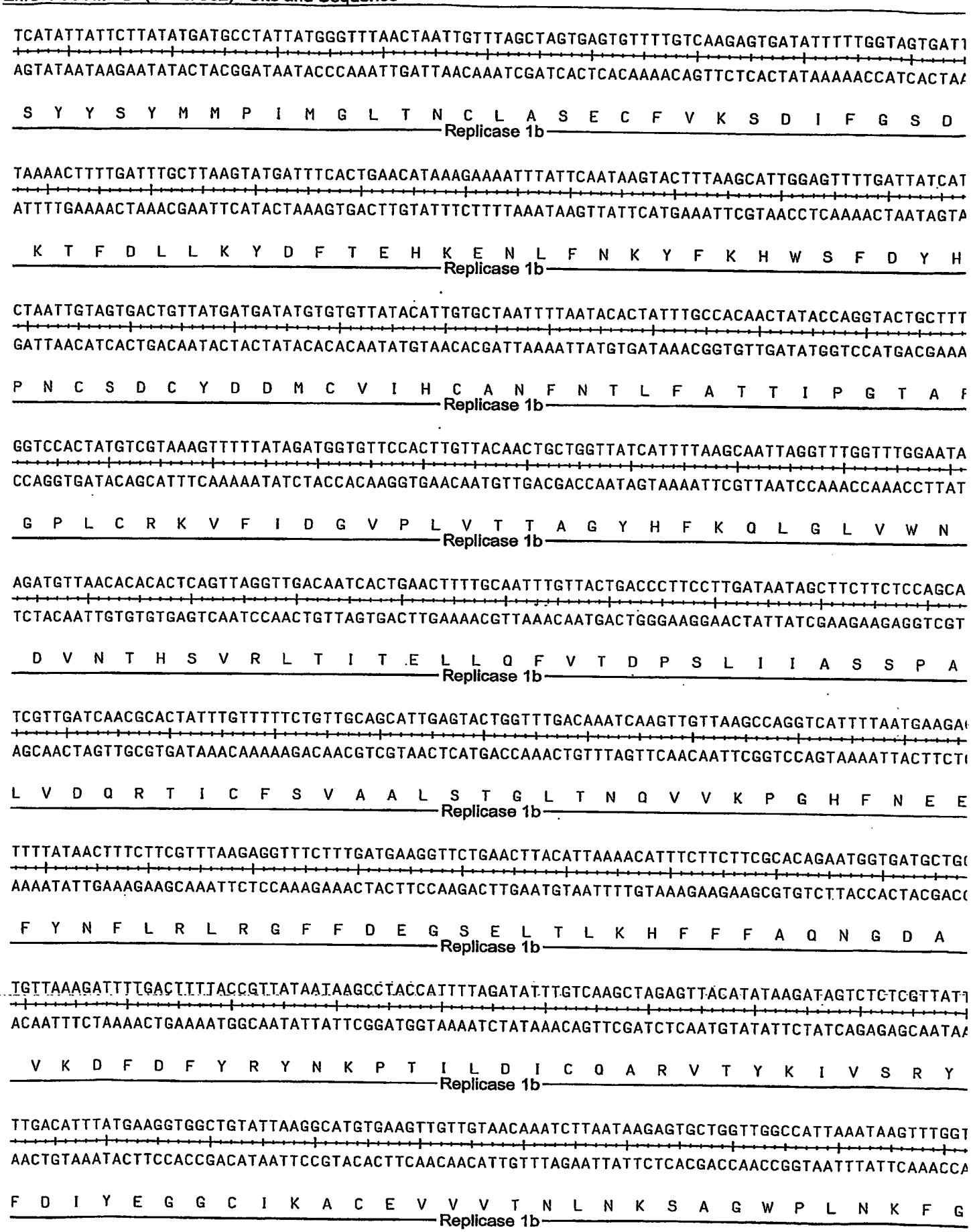
E V L V L T G C C D N S Y F D S K G W Y D P V E N E D I H R V
Replicase 1b

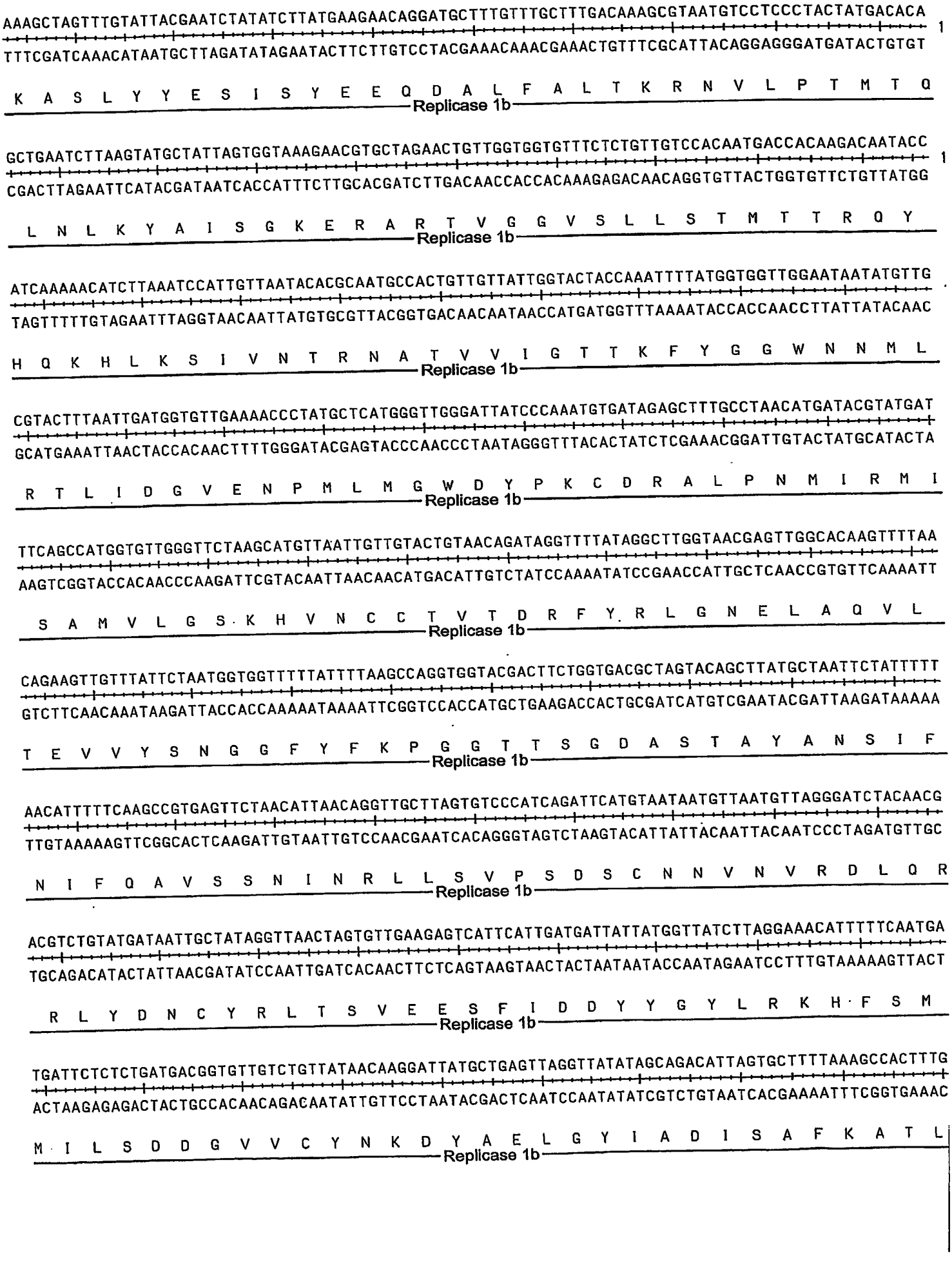
TTATGCATCTCTTGGCAAAATTGTAGCTAGAGCTATGCTTAAATGCGTTGCTCTATGTGATGCGATGGTTGCTAAAGGTGTTGTTGGTGTTT
AATACGTAGAGAACCGTTTTAACATCGATCTCGATACGAATTTACGCAACGAGATACACTACGCTACCAACGATTTCCACAACAACCACAAA

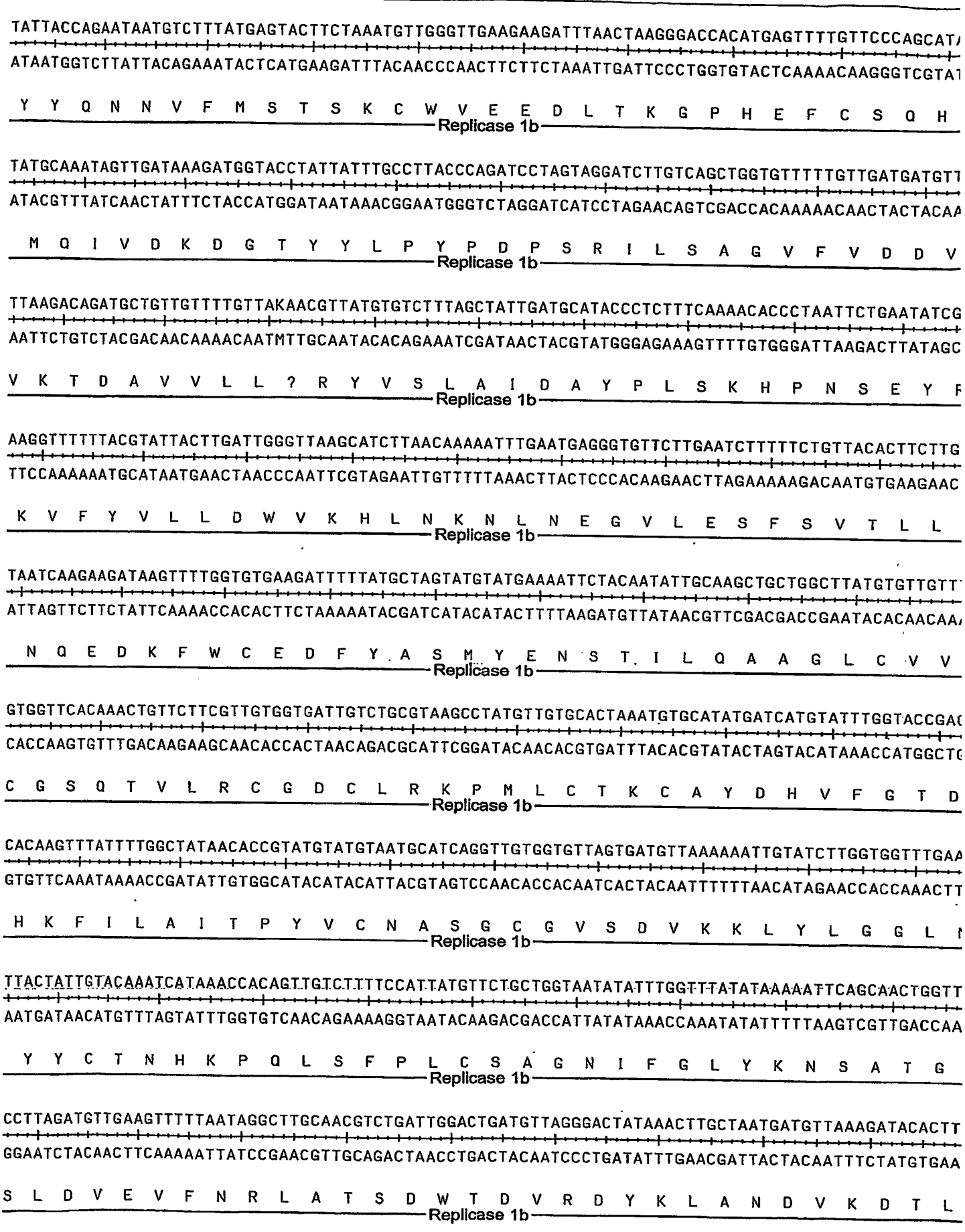
Y A S L G K I V A R A M L K C V A L C D A M V A K G V V G V
Replicase 1b

TAACATTAGATAACCAAGATCTTAATGGTAACTTTTATGATTTTGGTGATTTTGTTGTTAGCTTACCTAATATGGGTGTTCCCTGTTGTACA
ATTGTAATCTATTGGTTCTAGAATTACCATTGAAAATACTAAAACCACTAAAACAACAATCGAATGGATTATACCCACAAGGGACAACATGT

L T L D N Q D L N G N F Y D F G D F V V S L P N M G V P C C T
Replicase 1b


TCATATTATTCTTATATGATGCCTATTATGGGTTTAACTAATTGTTTAGCTAGTGAGTGTTTTGTCAAGAGTGATATTTTTGGTAGTGATT
AGTATAATAAGAATATACTACGGATAATACCCAAATTGATTAACAAATCGATCACTCACAAAACAGTTCTCACTATAAAAACCATCACTAA
S Y Y S Y M M P I M G L T N C L A S E C F V K S D I F G S D
Replicase 1b
TAAAACTTTTGATTTGCTTAAGTATGATTTCACTGAACATAAAGAAAATTTATTCAATAAGTACTTTAAGCATTGGAGTTTTGATTATCAT
ATTTTGAAAACTAAACGAATTCATACTAAAGTGACTTGTATTTCTTTTAAATAAGTTATTCATGAAATTCGTAACCTCAAAACTAATAGTA
K T F D L L K Y D F T E H K E N L F N K Y F K H W S F D Y H
Replicase 1b
CTAATTGTAGTGACTGTTATGATGATATGTGTGTTATACATTGTGCTAATTTTAATACACTATTTGCCACAACTATACCAGGTACTGCTTT
GATTAACATCACTGACAATACTACTATACACACAATATGTAACACGATTAAAATTATGTGATAAACGGTGTTGATATGGTCCATGACGAAA
P N C S D C Y D D M C V I H C A N F N T L F A T T I P G T A F
Replicase 1b
GGTCCACTATGTCGTAAAGTTTTTATAGATGGTGTTCCACTTGTTACAACTGCTGGTTATCATTTTAAGCAATTAGGTTTGGTTTGGAATA
CCAGGTGATACAGCATTTCAAAAATATCTACCACAAGGTGAACAATGTTGACGACCAATAGTAAAATTCGTTAATCCAAACCAAACCTTAT
G P L C R K V F I D G V P L V T T A G Y H F K Q L G L V W N
Replicase 1b
AGATGTTAACACACACTCAGTTAGGTTGACAATCACTGAACTTTTGCAATTTGTTACTGACCCTTCCTTGATAATAGCTTCTTCTCCAGCA
TCTACAATTGTGTGTGAGTCAATCCAACTGTTAGTGACTTGAAAACGTTAAACAATGACTGGGAAGGAACTATTATCGAAGAAGAGGTCGT
D V N T H S V R L T I T E L L Q F V T D P S L I I A S S P A
Replicase 1b
TCGTTGATCAACGCACTATTTGTTTTTCTGTTGCAGCATTGAGTACTGGTTTGACAAATCAAGTTGTTAAGCCAGGTCATTTTAATGAAGA
AGCAACTAGTTGCGTGATAAACAAAAAGACAACGTCGTAACTCATGACCAAACTGTTTAGTTCAACAATTCGGTCCAGTAAAATTACTTCT
L V D Q R T I C F S V A A L S T G L T N Q V V K P G H F N E E
Replicase 1b
TTTTATAACTTTCTTCGTTTAAGAGGTTTCTTTGATGAAGGTTCTGAACTTACATTAAAACATTTCTTCTTCGCACAGAATGGTGATGCTG
AAAATATTGAAAGAAGCAAATTCTCCAAAGAAACTACTTCCAAGACTTGAATGTAATTTTGTAAAGAAGAAGCGTGTCTTACCACTACGAC
F Y N F L R L R G F F D E G S E L T L K H F F F A Q N G D A
Replicase 1b
TGTTAAAGATTTTGACTTTTACCGTTATAATAAGCCTACCATTTTAGATATTTGTCAAGCTAGAGTTACATATAAGATAGTCTCTCGTTATT
ACAATTTCTAAAACTGAAAATGGCAATATTATTCGGATGGTAAAATCTATAAACAGTTCGATCTCAATGTATATTCTATCAGAGAGCAATAA
V K D F D F Y R Y N K P T I L D I C Q A R V T Y K I V S R Y
Replicase 1b
TTGACATTTATGAAGGTGGCTGTATTAAGGCATGTGAAGTTGTTGTAACAAATCTTAATAAGAGTGCTGGTTGGCCATTAAATAAGTTTGGT
AACTGTAAATACTTCCACCGACATAATTCCGTACACTTCAACAACATTGTTTAGAATTATTCTCACGACCAACCGGTAATTTATTCAAACCA
F D I Y E G G C I K A C E V V V T N L N K S A G W P L N K F G
Replicase 1b


AAAGCTAGTTTGTATTACGAATCTATATCTTATGAAGAACAGGATGCTTTGTTTGCTTTGACAAAGCGTAATGTCCTCCCTACTATGACACA 1
TTTCGATCAAACATAATGCTTAGATATAGAATACTTCTTGTCCTACGAAACAAACGAAACTGTTTCGCATTACAGGAGGGATGATACTGTGT
K A S L Y Y E S I S Y E E Q D A L F A L T K R N V L P T M T Q
Replicase 1b
GCTGAATCTTAAGTATGCTATTAGTGGTAAAGAACGTGCTAGAACTGTTGGTGGTGTTTCTCTGTTGTCCACAATGACCACAAGACAATACC 1
CGACTTAGAATTCATACGATAATCACCATTTCTTGCACGATCTTGACAACCACCACAAAGAGACAACAGGTGTTACTGGTGTTCTGTTATGG
L N L K Y A I S G K E R A R T V G G V S L L S T M T T R Q Y
Replicase 1b
ATCAAAAACATCTTAAATCCATTGTTAATACACGCAATGCCACTGTTGTTATTGGTACTACCAAATTTTATGGTGGTTGGAATAATATGTTG
TAGTTTTTGTAGAATTTAGGTAACAATTATGTGCGTTACGGTGACAACAATAACCATGATGGTTTAAAATACCACCAACCTTATTATACAAC
H Q K H L K S I V N T R N A T V V I G T T K F Y G G W N N M L
Replicase 1b
CGTACTTTAATTGATGGTGTTGAAAACCCTATGCTCATGGGTTGGGATTATCCCAAATGTGATAGAGCTTTGCCTAACATGATACGTATGAT
GCATGAAATTAACTACCACAACTTTTGGGATACGAGTACCCAACCCTAATAGGGTTTACACTATCTCGAAACGGATTGTACTATGCATACTA
R T L I D G V E N P M L M G W D Y P K C D R A L P N M I R M I
Replicase 1b
TTCAGCCATGGTGTTGGGTTCTAAGCATGTTAATTGTTGTACTGTAACAGATAGGTTTTATAGGCTTGGTAACGAGTTGGCACAAGTTTTAA
AAGTCGGTACCACAACCCAAGATTCGTACAATTAACAACATGACATTGTCTATCCAAAATATCCGAACCATTGCTCAACCGTGTTCAAAATT
S A M V L G S K H V N C C T V T D R F Y R L G N E L A Q V L
Replicase 1b
CAGAAGTTGTTTATTCTAATGGTGGTTTTTATTTTAAGCCAGGTGGTACGACTTCTGGTGACGCTAGTACAGCTTATGCTAATTCTATTTTT
GTCTTCAACAAATAAGATTACCACCAAAAATAAAATTCGGTCCACCATGCTGAAGACCACTGCGATCATGTCGAATACGATTAAGATAAAAA
T E V V Y S N G G F Y F K P G G T T S G D A S T A Y A N S I F
Replicase 1b
AACATTTTTCAAGCCGTGAGTTCTAACATTAACAGGTTGCTTAGTGTCCCATCAGATTCATGTAATAATGTTAATGTTAGGGATCTACAACG
TTGTAAAAAGTTCGGCACTCAAGATTGTAATTGTCCAACGAATCACAGGGTAGTCTAAGTACATTATTACAATTACAATCCCTAGATGTTGC
N I F Q A V S S N I N R L L S V P S D S C N N V N V R D L Q R
Replicase 1b
ACGTCTGTATGATAATTGCTATAGGTTAACTAGTGTTGAAGAGTCATTCATTGATGATTATTATGGTTATCTTAGGAAACATTTTTCAATGA
TGCAGACATACTATTAACGATATCCAATTGATCACAACTTCTCAGTAAGTAACTACTAATAATACCAATAGAATCCTTTGTAAAAAGTTACT
R L Y D N C Y R L T S V E E S F I D D Y Y G Y L R K H F S M
Replicase 1b
TGATTCTCTCTGATGACGGTGTTGTCTGTTATAACAAGGATTATGCTGAGTTAGGTTATATAGCAGACATTAGTGCTTTTAAAGCCACTTTG
ACTAAGAGAGACTACTGCCACAACAGACAATATTGTTCCTAATACGACTCAATCCAATATATCGTCTGTAATCACGAAAATTTCGGTGAAAC
M I L S D D G V V C Y N K D Y A E L G Y I A D I S A F K A T L
Replicase 1b


TATTACCAGAATAATGTCTTTATGAGTACTTCTAAATGTTGGGTTGAAGAAGATTTAACTAAGGGACCACATGAGTTTTGTTCCCAGCAT
ATAATGGTCTTATTACAGAAATACTCATGAAGATTTACAACCCAACTTCTTCTAAATTGATTCCCTGGTGTACTCAAAACAAGGGTCGTAT
Y Y Q N N V F M S T S K C W V E E D L T K G P H E F C S Q H
Replicase 1b
TATGCAAATAGTTGATAAAGATGGTACCTATTATTTGCCTTACCCAGATCCTAGTAGGATCTTGTCAGCTGGTGTTTTTGTTGATGATGTT
ATACGTTTATCAACTATTTCTACCATGGATAATAAACGGAATGGGTCTAGGATCATCCTAGAACAGTCGACCACAAAAACAACTACTACAA
M Q I V D K D G T Y Y L P Y P D P S R I L S A G V F V D D V
Replicase 1b
TTAAGACAGATGCTGTTGTTTTGTTAKAACGTTATGTGTCTTTAGCTATTGATGCATACCCTCTTTCAAAACACCCTAATTCTGAATATCG
AATTCTGTCTACGACAACAAAACAATMTTGCAATACACAGAAATCGATAACTACGTATGGGAGAAAGTTTTGTGGGATTAAGACTTATAGC
V K T D A V V L L ? R Y V S L A I D A Y P L S K H P N S E Y F
Replicase 1b
AAGGTTTTTTACGTATTACTTGATTGGGTTAAGCATCTTAACAAAAATTTGAATGAGGGTGTTCTTGAATCTTTTTCTGTTACACTTCTTG
TTCCAAAAAATGCATAATGAACTAACCCAATTCGTAGAATTGTTTTTAAACTTACTCCCACAAGAACTTAGAAAAAGACAATGTGAAGAAC
K V F Y V L L D W V K H L N K N L N E G V L E S F S V T L L
Replicase 1b
TAATCAAGAAGATAAGTTTTGGTGTGAAGATTTTTATGCTAGTATGTATGAAAATTCTACAATATTGCAAGCTGCTGGCTTATGTGTTGTT
ATTAGTTCTTCTATTCAAAACCACACTTCTAAAAATACGATCATACATACTTTTAAGATGTTATAACGTTCGACGACCGAATACACAACAA
N Q E D K F W C E D F Y A S M Y E N S T I L Q A A G L C V V
Replicase 1b
GTGGTTCACAAACTGTTCTTCGTTGTGGTGATTGTCTGCGTAAGCCTATGTTGTGCACTAAATGTGCATATGATCATGTATTTGGTACCGAC
CACCAAGTGTTTGACAAGAAGCAACACCACTAACAGACGCATTCGGATACAACACGTGATTTACACGTATACTAGTACATAAACCATGGCTG
C G S Q T V L R C G D C L R K P M L C T K C A Y D H V F G T D
Replicase 1b
CACAAGTTTATTTTGGCTATAACACCGTATGTATGTAATGCATCAGGTTGTGGTGTTAGTGATGTTAAAAAATTGTATCTTGGTGGTTTGAA
GTGTTCAAATAAAACCGATATTGTGGCATACATACATTACGTAGTCCAACACCACAATCACTACAATTTTTTAACATAGAACCACCAAACTT
H K F I L A I T P Y V C N A S G C G V S D V K K L Y L G G L
Replicase 1b
TTACTATTGTACAAATCATAAACCACAGTTGTCTTTTCCATTATGTTCTGCTGGTAATATATTTGGTTTATATAAAATTCAGCAACTGGTT
AATGATAACATGTTTAGTATTTGGTGTCAACAGAAAAGGTAATACAAGACGACCATTATATAAACCAAATATATTTTTAAGTCGTTGACCAA
Y Y C T N H K P Q L S F P L C S A G N I F G L Y K N S A T G
Replicase 1b
CCTTAGATGTTGAAGTTTTTAATAGGCTTGCAACGTCTGATTGGACTGATGTTAGGGACTATAAACTTGCTAATGATGTTAAAGATACACTT
GGAATCTACAACTTCAAAAATTATCCGAACGTTGCAGACTAACCTGACTACAATCCCTGATATTTGAACGATTACTACAATTTCTATGTGAA
S L D V E V F N R L A T S D W T D V R D Y K L A N D V K D T L
Replicase 1b

AGACTCTTTGCGGCTGAAACTATTAAAGCTAAAGAAGAGAGTGTTAAGTCTTCTTATGCTTTTGCAACTCTTAAAGAGGTTGTTGGACCTAA
TCTGAGAAACGCCGACTTTGATAATTTCGATTTCTTCTCTCACAATTCAGAAGAATACGAAAACGTTGAGAATTTCTCCAACAACCTGGATT

R L F A A E T I K A K E E S V K S S Y A F A T L K E V V G P K
Replicase 1b

AGAATTGCTTCTTAGTTGGGAAAGTGGTAAAGTTAAACCACCTTTGAATCGTAATTCTGTTTTCACCTGTTTTCAAATAAGTAAGGACTCAA
TCTTAACGAAGAATCAACCCTTTCACCATTTCAATTTGGTGGAAACTTAGCATTAAGACAAAAGTGGACAAAAGTTTATTCATTCCTGAGTT

E L L L S W E S G K V K P P L N R N S V F T C F Q I S K D S
Replicase 1b

AATTCCAAATAGGTGAGTTCATCTTTGAAAAGGTTGAATATGGTTCTGATACTGTTACGTATAAGTCTACTGTAACCACTAAGTTAGTTCCT
TTAAGGTTTATCCACTCAAGTAGAAACTTTTCCAACTTATACCAAGACTATGACAATGCATATTCAGATGACATTGGTGATTCAATCAAGGA

K F Q I G E F I F E K V E Y G S D T V T Y K S T V T T K L V P
Replicase 1b

GGTATGATTTTTGTCTTAACATCTCACAATGTTCAACCTTTACGTGCACCAACTATTGCAAACCAAGAGAAGTATTCTAGCATTTATAAATT
CCATACTAAAAACAGAATTGTAGAGTGTTACAAGTTGGAAATGCACGTGGTTGATAACGTTTGGTTCTCTTCATAAGATCGTAAATATTTAA

G M I F V L T S H N V Q P L R A P T I A N Q E K Y S S I Y K L
Replicase 1b

GCACCCTGCTTTTAATGTCAGTGATGCATATGCTAATTTGGTTCCATATTACCAACTTATTGGTAAACAAAAGATAACTACAATACAGGGTC
CGTGGGACGAAAATTACAGTCACTACGTATACGATTAAACCAAGGTATAATGGTTGAATAACCATTTGTTTTCTATTGATGTTATGTCCCAG

H P A F N V S D A Y A N L V P Y Y Q L I G K Q K I T T I Q G
Replicase 1b

CTCCTGGTAGTGGTAAGTCACATTGTTCCATTGGACTTGGATTGTACTATCCAGGTGCGCGTATTGTTTTTGTTGCTTGTGCCCATGCTGCT
GAGGACCATCACCATTCAGTGTAACAAGGTAACCTGAACCTAACATGATAGGTCCACGCGCATAACAAAAACAACGAACACGGGTACGACGA

P P G S G K S H C S I G L G L Y Y P G A R I V F V A C A H A A
Replicase 1b

GTTGATTCCTTATGTGCAAAAGCTATGACTGTTTATAGCATTGATAAGTGTACTAGGATTATACCTGCAAGAGCTCGGGTTGAGTGTTATAG
CAACTAAGGAATACACGTTTTCGATACTGACAAATATCGTAACTATTCACATGATCCTAATATGGACGTTCTCGAGCCCAACTCACAATATC

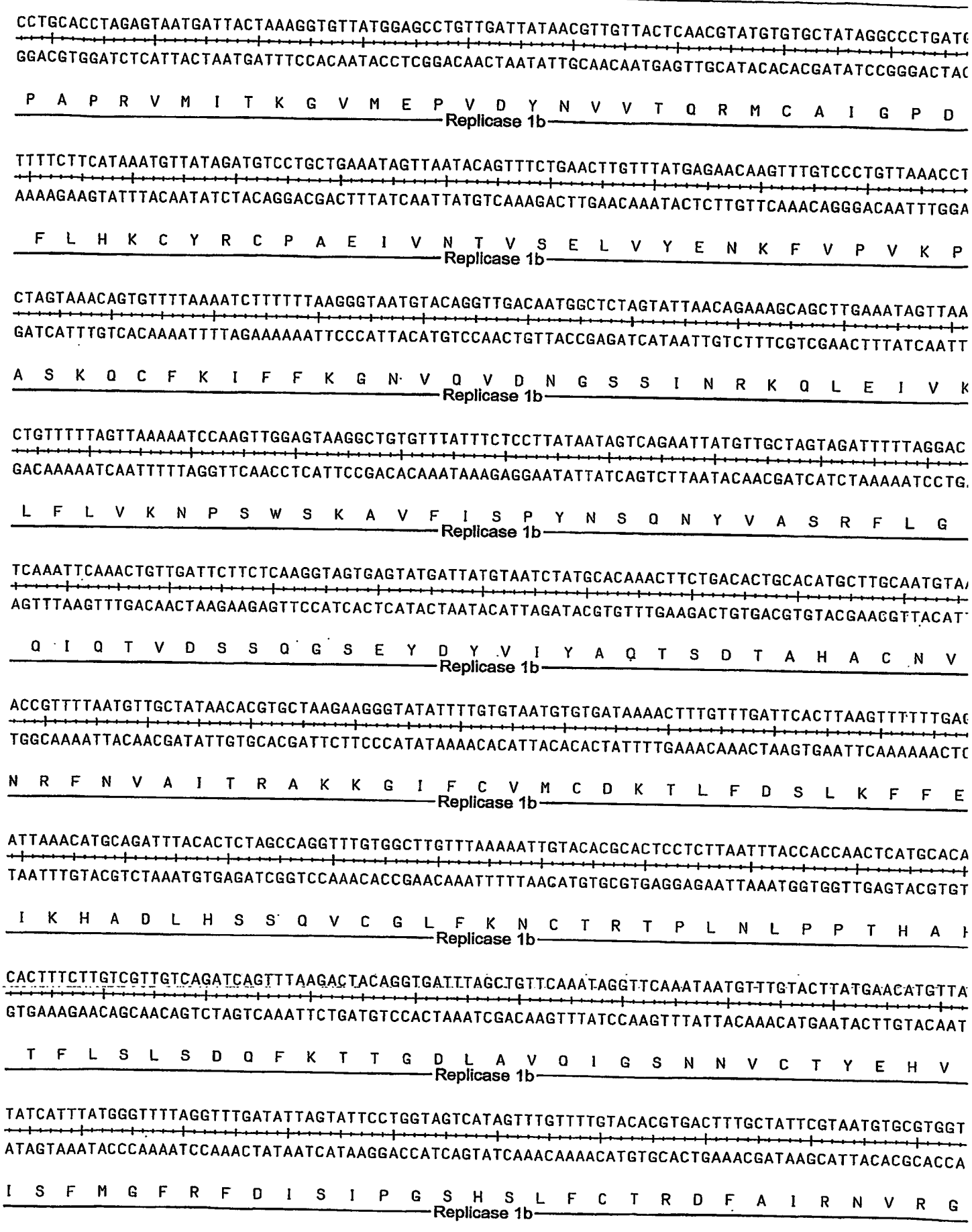
V D S L C A K A M T V Y S I D K C T R I I P A R A R V E C Y S
Replicase 1b

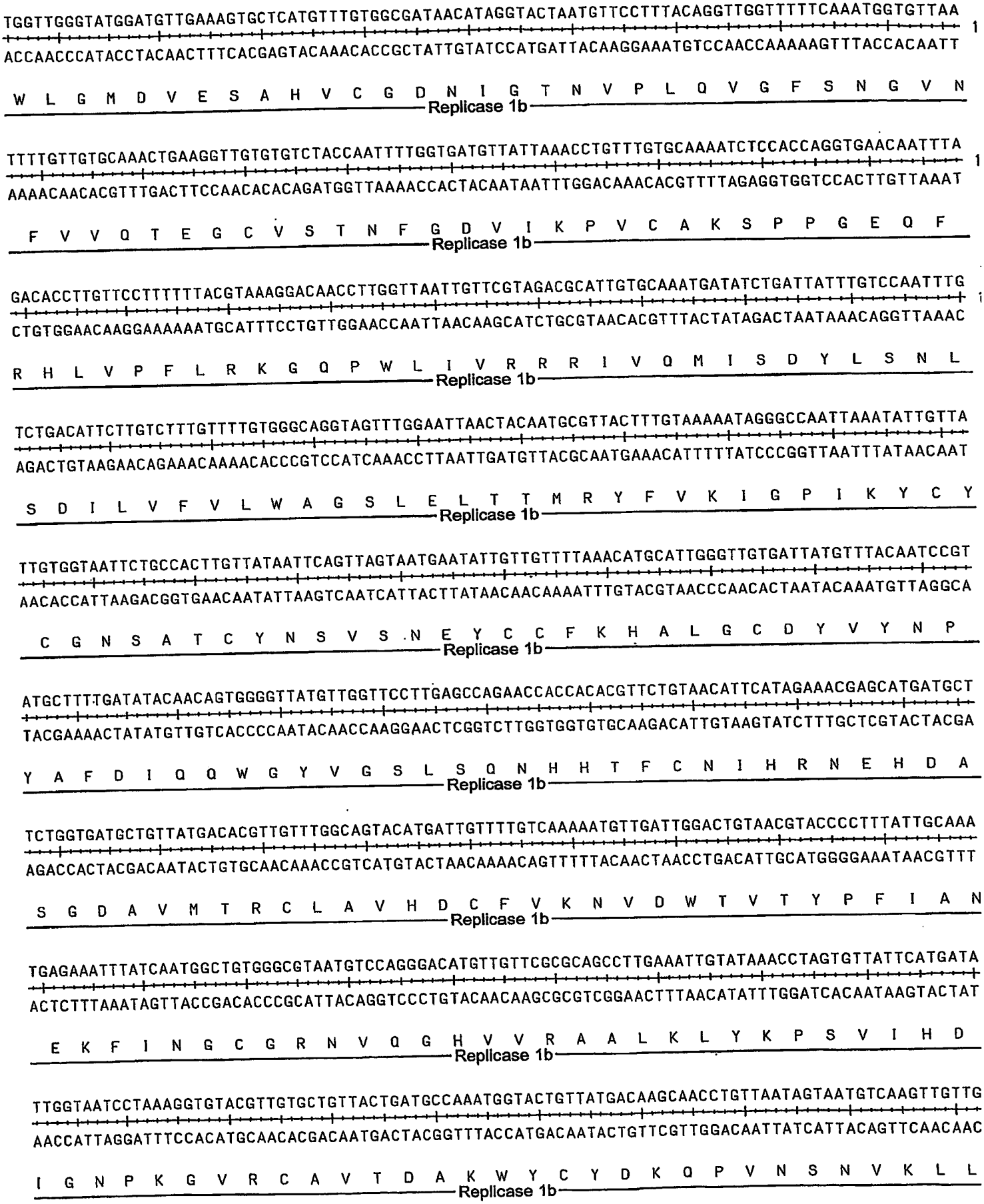
TGGCTTTAAACCAAATAACACTAGTGCACAATACATATTTAGCACTGTTAACGCATTACCTGAGTGTAATGCTGATATTGTTGTTGTAGATG
ACCGAAATTTGGTTTATTGTGATCACGTGTTATGTATAAATCGTGACAATTGCGTAATGGACTCACATTACGACTATAACAACAACATCTAC

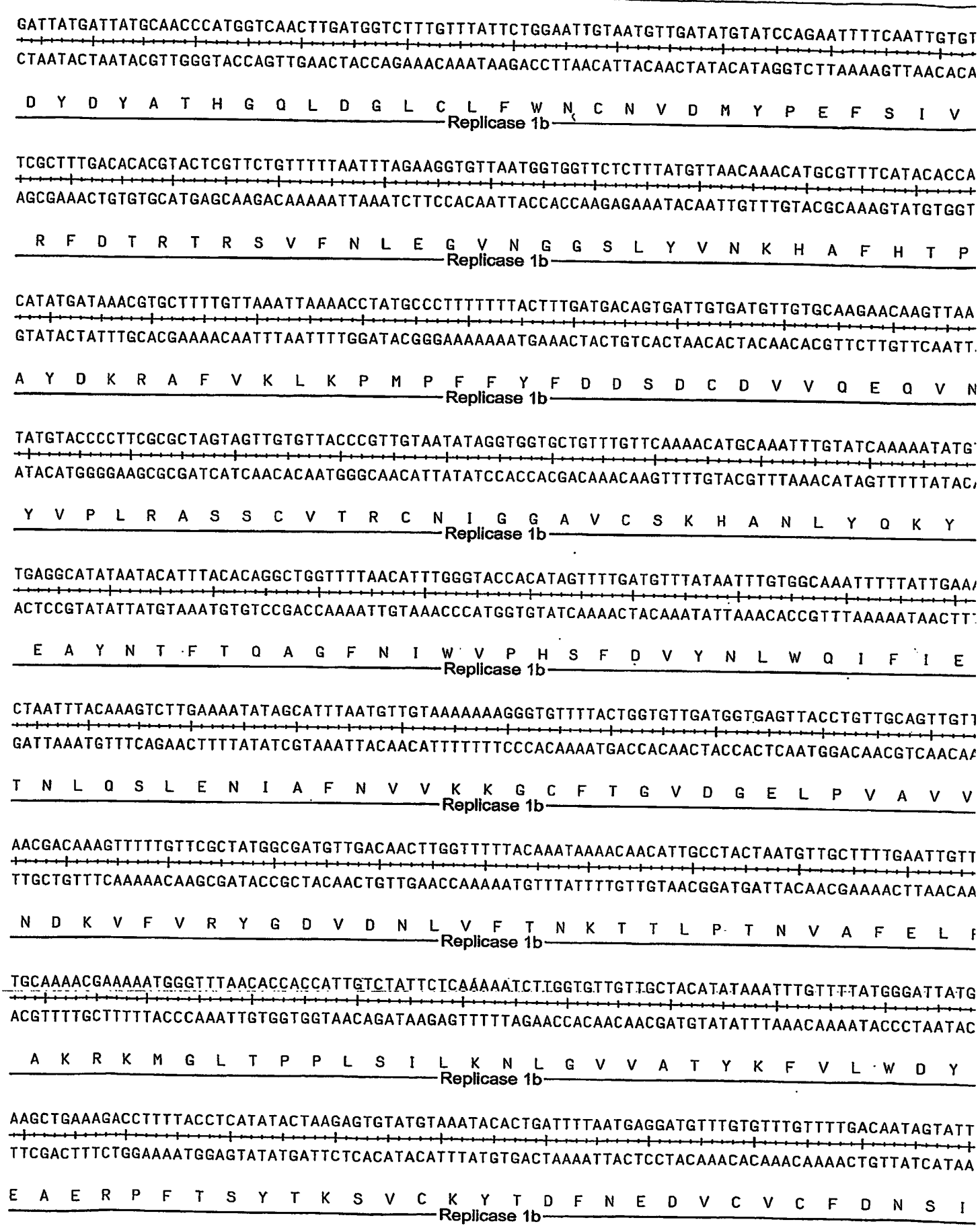
G F K P N N T S A Q Y I F S T V N A L P E C N A D I V V V D
Replicase 1b

AAGTTTCAATGTGTACAAATTATGACCTTTCTGTTATTAATCAGCGTTTATCATATAAACATATTGTTTATGTTGGTGATCCACAACAACTT
TTCAAAGTTACACATGTTTAATACTGGAAAGACAATAATTAGTCGCAAATAGTATATTTGTATAACAAATACAACCACTAGGTGTTGTTGAA

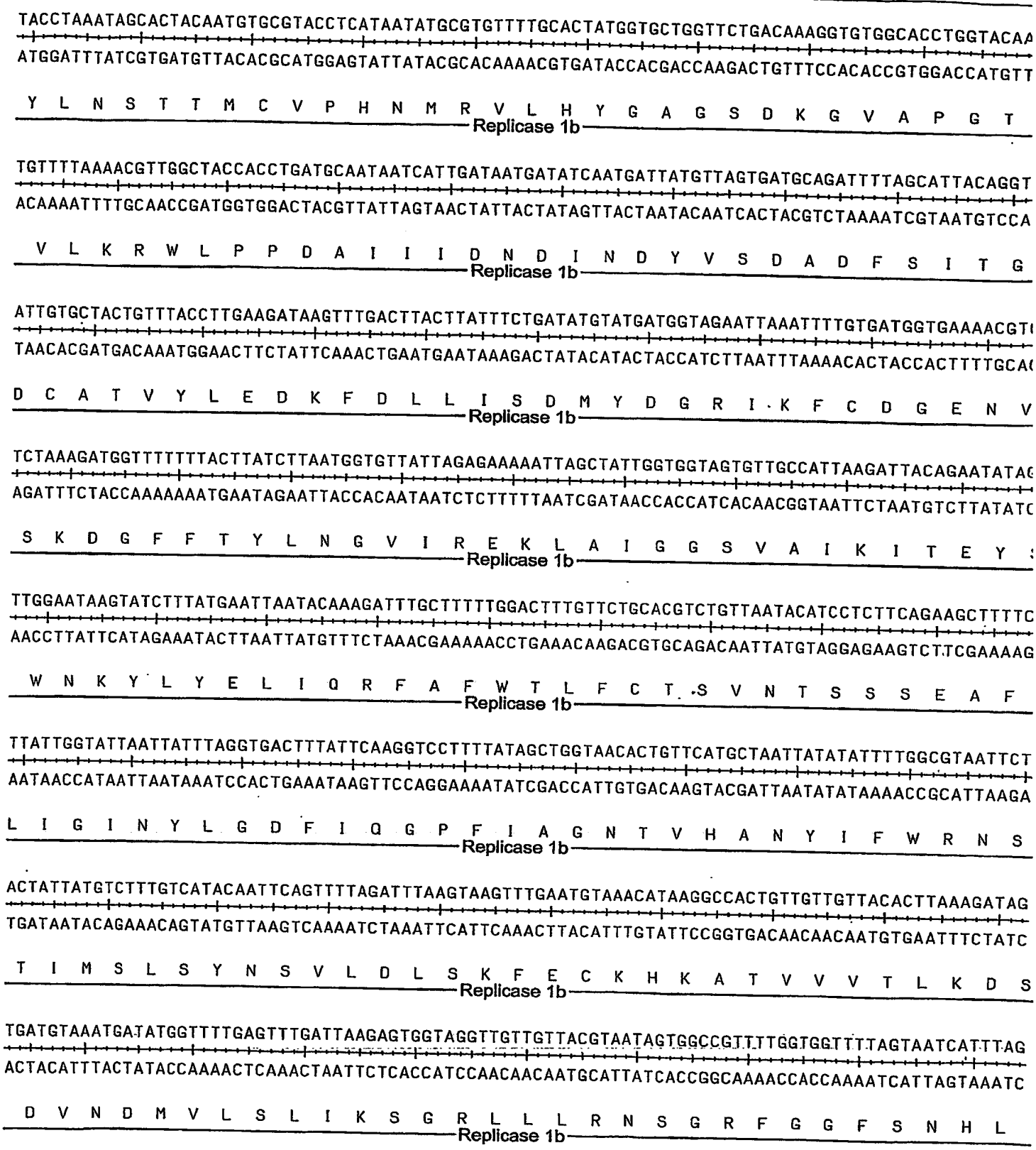
E V S M C T N Y D L S V I N Q R L S Y K H I V Y V G D P Q Q L
Replicase 1b

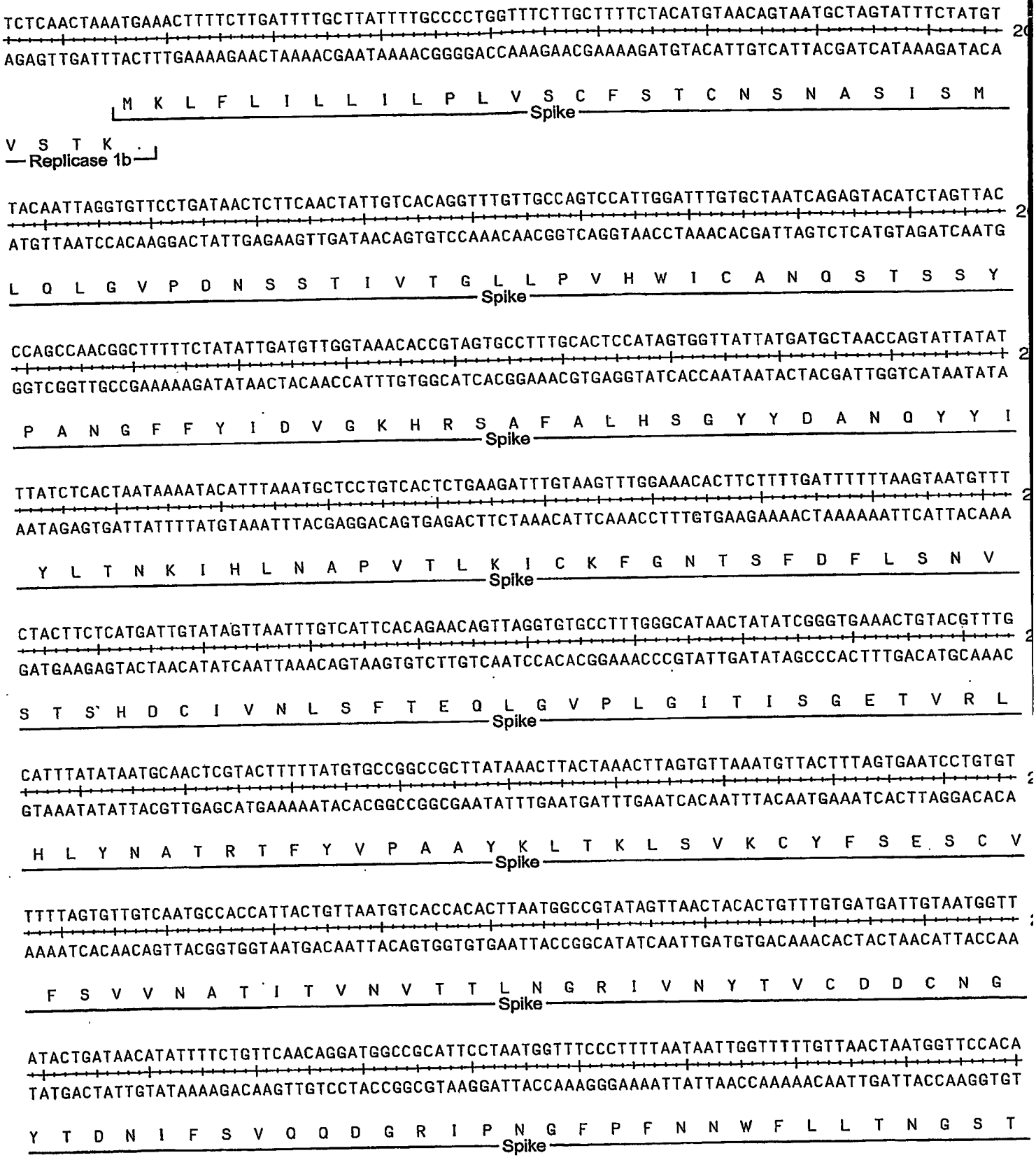

CCTGCACCTAGAGTAATGATTACTAAAGGTGTTATGGAGCCTGTTGATTATAACGTTGTTACTCAACGTATGTGTGCTATAGGCCCTGATG
GGACGTGGATCTCATTACTAATGATTTCCACAATACCTCGGACAACTAATATTGCAACAATGAGTTGCATACACACGATATCCGGGACTAC
P A P R V M I T K G V M E P V D Y N V V T Q R M C A I G P D
Replicase 1b
TTTTCTTCATAAATGTTATAGATGTCCTGCTGAAATAGTTAATACAGTTTCTGAACTTGTTTATGAGAACAAGTTTGTCCCTGTTAAACCT
AAAAGAAGTATTTACAATATCTACAGGACGACTTTATCAATTATGTCAAAGACTTGAACAAATACTCTTGTTCAAACAGGGACAATTTGGA
F L H K C Y R C P A E I V N T V S E L V Y E N K F V P V K P
Replicase 1b
CTAGTAAACAGTGTTTTAAAATCTTTTTTAAGGGTAATGTACAGGTTGACAATGGCTCTAGTATTAACAGAAAGCAGCTTGAAATAGTTAA
GATCATTTGTCACAAAATTTTAGAAAAAATTCCCATTACATGTCCAACTGTTACCGAGATCATAATTGTCTTTCGTCGAACTTTATCAATT
A S K Q C F K I F F K G N V Q V D N G S S I N R K Q L E I V K
Replicase 1b
CTGTTTTTAGTTAAAAATCCAAGTTGGAGTAAGGCTGTGTTTATTTCTCCTTATAATAGTCAGAATTATGTTGCTAGTAGATTTTTAGGAC
GACAAAAATCAATTTTTAGGTTCAACCTCATTCCGACACAAATAAAGAGGAATATTATCAGTCTTAATACAACGATCATCTAAAAATCCTG
L F L V K N P S W S K A V F I S P Y N S Q N Y V A S R F L G
Replicase 1b
TCAAATTCAAACTGTTGATTCTTCTCAAGGTAGTGAGTATGATTATGTAATCTATGCACAAACTTCTGACACTGCACATGCTTGCAATGTA
AGTTTAAGTTTGACAACTAAGAAGAGTTCCATCACTCATACTAATACATTAGATACGTGTTTGAAGACTGTGACGTGTACGAAGGTTACAT
Q I Q T V D S S Q G S E Y D Y V I Y A Q T S D T A H A C N V
Replicase 1b
ACCGTTTTAATGTTGCTATAACACGTGCTAAGAAGGGTATATTTTGTGTAATGTGTGATAAAACTTTGTTTGATTCACTTAAGTTTTTTGAG
TGGCAAAATTACAACGATATTGTGCACGATTCTTCCCATATAAAACACATTACACACTATTTTGAAACAAACTAAGTGAATTCAAAAAACTC
N R F N V A I T R A K K G I F C V M C D K T L F D S L K F F E
Replicase 1b
ATTAAACATGCAGATTTACACTCTAGCCAGGTTTGTGGCTTGTTTAAAAATTGTACACGCACTCCTCTTAATTTACCACCAACTCATGCACA
TAATTTGTACGTCTAAATGTGAGATCGGTCCAAACACCGAACAAATTTTTAACATGTGCGTGAGGAGAATTAAATGGTGGTTGAGTACGTGT
I K H A D L H S S Q V C G L F K N C T R T P L N L P P T H A
Replicase 1b
CACTTTCTTGTCGTTGTCAGATCAGTTTAAGACTACAGGTGATTTAGCTGTTCAAATAGGTTCAAATAATGTTTGTACTTATGAACATGTTA
GTGAAAGAACAGCAACAGTCTAGTCAAATTCTGATGTCCACTAAATCGACAAGTTTATCCAAGTTTATTACAAACATGAATACTTGTACAAT
T F L S L S D Q F K T T G D L A V Q I G S N N V C T Y E H V
Replicase 1b
TATCATTTATGGGTTTTAGGTTTGATATTAGTATTCCTGGTAGTCATAGTTTGTTTTGTACACGTGACTTTGCTATTCGTAATGTGCGTGGT
ATAGTAAATACCCAAAATCCAAACTATAATCATAAGGACCATCAGTATCAAACAAAACATGTGCACTGAAACGATAAGCATTACACGCACCA
I S F M G F R F D I S I P G S H S L F C T R D F A I R N V R G
Replicase 1b


TGGTTGGGTATGGATGTTGAAAGTGCTCATGTTTGTGGCGATAACATAGGTACTAATGTTCCTTTACAGGTTGGTTTTTCAAATGGTGTTAA
ACCAACCCATACCTACAACTTTCACGAGTACAAACACCGCTATTGTATCCATGATTACAAGGAAATGTCCAACCAAAAAGTTTACCACAATT
W L G M D V E S A H V C G D N I G T N V P L Q V G F S N G V N
Replicase 1b
TTTTGTTGTGCAAACTGAAGGTTGTGTGTCTACCAATTTTGGTGATGTTATTAAACCTGTTTGTGCAAAATCTCCACCAGGTGAACAATTTA
AAAACAACACGTTTGACTTCCAACACACAGATGGTTAAAACCACTACAATAATTTGGACAAACACGTTTTAGAGGTGGTCCACTTGTTAAAT
F V V Q T E G C V S T N F G D V I K P V C A K S P P G E Q F
Replicase 1b
GACACCTTGTTCCTTTTTTACGTAAAGGACAACCTTGGTTAATTGTTCGTAGACGCATTGTGCAAATGATATCTGATTATTTGTCCAATTTG
CTGTGGAACAAGGAAAAAATGCATTTCCTGTTGGAACCAATTAACAAGCATCTGCGTAACACGTTTACTATAGACTAATAAACAGGTTAAAC
R H L V P F L R K G Q P W L I V R R R I V Q M I S D Y L S N L
Replicase 1b
TCTGACATTCTTGTCTTTGTTTTGTGGGCAGGTAGTTTGGAATTAACTACAATGCGTTACTTTGTAAAAATAGGGCCAATTAAATATTGTTA
AGACTGTAAGAACAGAAACAAAACACCCGTCCATCAAACCTTAATTGATGTTACGCAATGAAACATTTTTATCCCGGTTAATTTATAACAAT
S D I L V F V L W A G S L E L T T M R Y F V K I G P I K Y C Y
Replicase 1b
TTGTGGTAATTCTGCCACTTGTTATAATTCAGTTAGTAATGAATATTGTTGTTTTAAACATGCATTGGGTTGTGATTATGTTTACAATCCGT
AACACCATTAAGACGGTGAACAATATTAAGTCAATCATTACTTATAACAACAAAATTTGTACGTAACCCAACACTAATACAAATGTTAGGCA
C G N S A T C Y N S V S N E Y C C F K H A L G C D Y V Y N P
Replicase 1b
ATGCTTTTGATATACAACAGTGGGGTTATGTTGGTTCCTTGAGCCAGAACCACCACACGTTCTGTAACATTCATAGAAACGAGCATGATGCT
TACGAAAACTATATGTTGTCACCCCAATACAACCAAGGAACTCGGTCTTGGTGGTGTGCAAGACATTGTAAGTATCTTTGCTCGTACTACGA
Y A F D I Q Q W G Y V G S L S Q N H H T F C N I H R N E H D A
Replicase 1b
TCTGGTGATGCTGTTATGACACGTTGTTTGGCAGTACATGATTGTTTTGTCAAAAATGTTGATTGGACTGTAACGTACCCCTTTATTGCAAA
AGACCACTACGACAATACTGTGCAACAAACCGTCATGTACTAACAAAACAGTTTTTACAACTAACCTGACATTGCATGGGGAAATAACGTTT
S G D A V M T R C L A V H D C F V K N V D W T V T Y P F I A N
Replicase 1b
TGAGAAATTTATCAATGGCTGTGGGCGTAATGTCCAGGGACATGTTGTTCGCGCAGCCTTGAAATTGTATAAACCTAGTGTTATTCATGATA
ACTCTTTAAATAGTTACCGACACCCGCATTACAGGTCCCTGTACAACAAGCGCGTCGGAACTTTAACATATTTGGATCACAATAAGTACTAT
E K F I N G C G R N V Q G H V V R A A L K L Y K P S V I H D
Replicase 1b
TTGGTAATCCTAAAGGTGTACGTTGTGCTGTTACTGATGCCAAATGGTACTGTTATGACAAGCAACCTGTTAATAGTAATGTCAAGTTGTTG
AACCATTAGGATTTCCACATGCAACACGACAATGACTACGGTTTACCATGACAATACTGTTCGTTGGACAATTATCATTACAGTTCAACAAC
I G N P K G V R C A V T D A K W Y C Y D K Q P V N S N V K L L
Replicase 1b


GATTATGATTATGCAACCCATGGTCAACTTGATGGTCTTTGTTTATTCTGGAATTGTAATGTTGATATGTATCCAGAATTTTCAATTGTGT
CTAATACTAATACGTTGGGTACCAGTTGAACTACCAGAAACAAATAAGACCTTAACATTACAACTATACATAGGTCTTAAAAGTTAACACA
D Y D Y A T H G Q L D G L C L F W N C N V D M Y P E F S I V
Replicase 1b
TCGCTTTGACACACGTACTCGTTCTGTTTTTAATTTAGAAGGTGTTAATGGTGGTTCTCTTTATGTTAACAAACATGCGTTTCATACACCA
AGCGAAACTGTGTGCATGAGCAAGACAAAAATTAAATCTTCCACAATTACCACCAAGAGAAATACAATTGTTTGTACGCAAAGTATGTGGT
R F D T R T R S V F N L E G V N G G S L Y V N K H A F H T P
Replicase 1b
CATATGATAAACGTGCTTTTGTTAAATTAAAACCTATGCCCTTTTTTTACTTTGATGACAGTGATTGTGATGTTGTGCAAGAACAAGTTAA
GTATACTATTTGCACGAAAACAATTTAATTTTGGATACGGGAAAAAAATGAAACTACTGTCACTAACACTACAACACGTTCTTGTTCAATT
A Y D K R A F V K L K P M P F F Y F D D S D C D V V Q E Q V N
Replicase 1b
TATGTACCCCTTCGCGCTAGTAGTTGTGTTACCCGTTGTAATATAGGTGGTGCTGTTTGTTCAAAACATGCAAATTTGTATCAAAAATATG
ATACATGGGGAAGCGCGATCATCAACACAATGGGCAACATTATATCCACCACGACAAACAAGTTTTGTACGTTTAAACATAGTTTTTATAC
Y V P L R A S S C V T R C N I G G A V C S K H A N L Y Q K Y
Replicase 1b
TGAGGCATATAATACATTTACACAGGCTGGTTTTAACATTTGGGTACCACATAGTTTTGATGTTTATAATTTGTGGCAAATTTTTATTGAA
ACTCCGTATATTATGTAAATGTGTCCGACCAAAATTGTAAACCCATGGTGTATCAAAACTACAAATATTAAACACCGTTTAAAAATAACTT
E A Y N T F T Q A G F N I W V P H S F D V Y N L W Q I F I E
Replicase 1b
CTAATTTACAAAGTCTTGAAAATATAGCATTTAATGTTGTAAAAAAAGGGTGTTTTACTGGTGTTGATGGTGAGTTACCTGTTGCAGTTGT
GATTAAATGTTTCAGAACTTTTATATCGTAAATTACAACATTTTTTTCCCACAAAATGACCACAACTACCACTCAATGGACAACGTCAACA
T N L Q S L E N I A F N V V K K G C F T G V D G E L P V A V V
Replicase 1b
AACGACAAAGTTTTTGTTCGCTATGGCGATGTTGACAACTTGGTTTTTACAAATAAAACAACATTGCCTACTAATGTTGCTTTTGAATTGTT
TTGCTGTTTCAAAAACAAGCGATACCGCTACAACTGTTGAACCAAAAATGTTTATTTTGTTGTAACGGATGATTACAACGAAAACTTAACAA
N D K V F V R Y G D V D N L V F T N K T T L P T N V A F E L
Replicase 1b
TGCAAAACGAAAAATGGGTTTAACACCACCATTGTCTATTCTCAAAAATCTTGGTGTTGTTGCTACATATAAATTTGTTTTATGGGATTATG
ACGTTTTGCTTTTTACCCAAATTGTGGTGGTAACAGATAAGAGTTTTTAGAACCACAACAACGATGTATATTTAAACAAAATACCCTAATAC
A K R K M G L T P P L S I L K N L G V V A T Y K F V L W D Y
Replicase 1b
AAGCTGAAAGACCTTTTACCTCATATACTAAGAGTGTATGTAAATACACTGATTTTAATGAGGATGTTTGTGTTTGTTTTTGACAATAGTATT
TTCGACTTTCTGGAAAATGGAGTATATGATTCTCACATACATTTATGTGACTAAAATTACTCCTACAAACACAAACAAAACTGTTATCATAA
E A E R P F T S Y T K S V C K Y T D F N E D V C V C F D N S I
Replicase 1b

```
CAGGGTTCGTATGAGCGTTTTACGCTTACTACGAACGCTGTTTTATTTTCTACTGTTGTCATTAAAAATTTAACACCTATAAAGTTGAATTT
GTCCCAAGCATACTCGCAAAATGCGAATGATGCTTGCGACAAAATAAAAGATGACAACAGTAATTTTTAAATTGTGGATATTTCAACTTAAA

Q G S Y E R F T L T T N A V L F S T V V I K N L T P I K L N F
                    Replicase 1b

TGGTATGTTGAATGGTATGCCAGTTTCTTCTATTAAGAGTGATAAAGGTGTTGAAAAATTAGTTAATTGGTACACATATGTTCGTAAAAATG
ACCATACAACTTACCATACGGTCAAAGAAGATAATTCTCACTATTTCCACAACTTTTTAATCAATTAACCATGTGTATACAAGCATTTTTAC

G M L N G M P V S S I K S D K G V E K L V N W Y T Y V R K N
                    Replicase 1b

GTCAATTTCAAGATCATTATGATGGTTTTTACACTCAAGGTAGGAATTTATCAGACTTTACACCAAGAAGTGATATGGAGTATGATTTTCTT
CAGTTAAAGTTCTAGTAATACTACCAAAAATGTGAGTTCCATCCTTAAATAGTCTGAAATGTGGTTCTTCACTATACCTCATACTAAAAGAA

G Q F Q D H Y D G F Y T Q G R N L S D F T P R S D M E Y D F L
                    Replicase 1b

AACATGGATATGGGTGTTTTTATTAATAAATATGGTCTTGAGGATTTTAATTTTGAACATGTTGTATATGGTGATGTTTCAAAAACTACATT
TTGTACCTATACCCACAAAAATAATTATTTATACCAGAACTCCTAAAATTAAAACTTGTACAACATATACCACTACAAAGTTTTTGATGTAA

N M D M G V F I N K Y G L E D F N F E H V V Y G D V S K T T L
                    Replicase 1b

AGGAGGTCTTCATTTGTTGATATCACAGTTTAGGCTTAGTAAAATGGGTGTTTTGAAAGCTGATGATTTTGTCACTGCTTCTGACACAACTT
TCCTCCAGAAGTAAACAACTATAGTGTCAAATCCGAATCATTTTACCCACAAAACTTTCGACTACTAAAACAGTGACGAAGACTGTGTTGAA

G G L H L L I S Q F R L S K M G V L K A D D F V T A S D T T
                    Replicase 1b

TGAGGTGCTGTACTGTTACTTATCTTAATGAACTTAGTTCAAAAGTTGTTTGTACTTATATGGATTTGTTGTTGGACGACTTTGTTACTATA
ACTCCACGACATGACAATGAATAGAATTACTTGAATCAAGTTTTCAACAAACATGAATATACCTAAACAACAACCTGCTGAAACAATGATAT

L R C C T V T Y L N E L S S K V V C T Y M D L L L D D F V T I
                    Replicase 1b

CTAAAGAGTTTAGATCTTGGTGTAATATCTAAAGTTCATGAAGTTATTATAGATAATAAACCTTATAGGTGGATGTTGTGGTGTAAAGATAA
GATTTCTCAAATCTAGAACCACATTATAGATTTCAAGTACTTCAATAATATCTATTATTTGGAATATCCACCTACAACACCACATTTCTATT

L K S L D L G V I S K V H E V I I D N K P Y R W M L W C K D N
                    Replicase 1b

CCACTTGTCGACTTTTTATCCACAGTTGCAGTCTGCTGAATGGAAGTGTGGTTATGCTATGCCACAAATTTATAAGCTTCAACGTATGTGTT
GGTGAACAGCTGAAAAATAGGTGTCAACGTCAGACGACTTACCTTCACACCAATACGATACGGTGTTTAAATATTCGAAGTTGCATACACAA

H L S T F Y P Q L Q S A E W K C G Y A M P Q I Y K L Q R M C
                    Replicase 1b

TGGAACCTTGTAATTTATATAATTATGGTGCTGGTATTAAGTTGCCTAGTGGTATAATGTTAAATGTTGTTAAATACACTCAGCTTTGTCAA
ACCTTGGAACATTAAATATATTAATACCACGACCATAATTCAACGGATCACCATATTACAATTTACAACAATTTATGTGAGTCGAAACAGTT

L E P C N L Y N Y G A G I K L P S G I M L N V V K Y T Q L C Q
                    Replicase 1b
```


TACCTAAATAGCACTACAATGTGCGTACCTCATAATATGCGTGTTTTGCACTATGGTGCTGGTTCTGACAAAGGTGTGGCACCTGGTACAA
ATGGATTTATCGTGATGTTACACGCATGGAGTATTATACGCACAAAACGTGATACCACGACCAAGACTGTTTCCACACCGTGGACCATGTT
Y L N S T T M C V P H N M R V L H Y G A G S D K G V A P G T
Replicase 1b
TGTTTTAAAACGTTGGCTACCACCTGATGCAATAATCATTGATAATGATATCAATGATTATGTTAGTGATGCAGATTTTAGCATTACAGGT
ACAAAATTTTGCAACCGATGGTGGACTACGTTATTAGTAACTATTACTATAGTTACTAATACAATCACTACGTCTAAAATCGTAATGTCCA
V L K R W L P P D A I I I D N D I N D Y V S D A D F S I T G
Replicase 1b
ATTGTGCTACTGTTTACCTTGAAGATAAGTTTGACTTACTTATTTCTGATATGTATGATGGTAGAATTAAATTTTGTGATGGTGAAAACGT
TAACACGATGACAAATGGAACTTCTATTCAAACTGAATGAATAAAGACTATACATACTACCATCTTAATTTAAAACACTACCACTTTTGCA
D C A T V Y L E D K F D L L I S D M Y D G R I K F C D G E N V
Replicase 1b
TCTAAAGATGGTTTTTTTACTTATCTTAATGGTGTTATTAGAGAAAAATTAGCTATTGGTGGTAGTGTTGCCATTAAGATTACAGAATATA
AGATTTCTACCAAAAAAATGAATAGAATTACCACAATAATCTCTTTTTAATCGATAACCACCATCACAACGGTAATTCTAATGTCTTATATC
S K D G F F T Y L N G V I R E K L A I G G S V A I K I T E Y
Replicase 1b
TTGGAATAAGTATCTTTATGAATTAATACAAAGATTTGCTTTTTGGACTTTGTTCTGCACGTCTGTTAATACATCCTCTTCAGAAGCTTTTC
AACCTTATTCATAGAAATACTTAATTATGTTTCTAAACGAAAAACCTGAAACAAGACGTGCAGACAATTATGTAGGAGAAGTCTTCGAAAAG
W N K Y L Y E L I Q R F A F W T L F C T S V N T S S S E A F
Replicase 1b
TTATTGGTATTAATTATTTAGGTGACTTTATTCAAGGTCCTTTTATAGCTGGTAACACTGTTCATGCTAATTATATATTTTGGCGTAATTCT
AATAACCATAATTAATAAATCCACTGAAATAAGTTCCAGGAAAATATCGACCATTGTGACAAGTACGATTAATATATAAAACCGCATTAAGA
L I G I N Y L G D F I Q G P F I A G N T V H A N Y I F W R N S
Replicase 1b
ACTATTATGTCTTTGTCATACAATTCAGTTTTAGATTTAAGTAAGTTTGAATGTAAACATAAGGCCACTGTTGTTGTTACACTTAAAGATAG
TGATAATACAGAAACAGTATGTTAAGTCAAAATCTAAATTCATTCAAACTTACATTTGTATTCCGGTGACAACAACAATGTGAATTTCTATC
T I M S L S Y N S V L D L S K F E C K H K A T V V V T L K D S
Replicase 1b
TGATGTAAATGATATGGTTTTGAGTTTGATTAAGAGTGGTAGGTTGTTGTTACGTAATAGTGGCCGTTTTGGTGGTTTTAGTAATCATTTAG
ACTACATTTACTATACCAAAACTCAAACTAATTCTCACCATCCAACAACAATGCATTATCACCGGCAAAACCACCAAAATCATTAGTAAATC
D V N D M V L S L I K S G R L L L R N S G R F G G F S N H L
Replicase 1b


TCTCAACTAAATGAAACTTTTCTTGATTTTGCTTATTTTGCCCCTGGTTTCTTGCTTTTCTACATGTAACAGTAATGCTAGTATTTCTATGT
AGAGTTGATTTACTTTGAAAAGAACTAAAACGAATAAAACGGGGACCAAAGAACGAAAAGATGTACATTGTCATTACGATCATAAAGATACA
M K L F L I L L I L P L V S C F S T C N S N A S I S M
Spike
V S T K .
Replicase 1b
TACAATTAGGTGTTCCTGATAACTCTTCAACTATTGTCACAGGTTTGTTGCCAGTCCATTGGATTTGTGCTAATCAGAGTACATCTAGTTAC
ATGTTAATCCACAAGGACTATTGAGAAGTTGATAACAGTGTCCAAACAACGGTCAGGTAACCTAAACACGATTAGTCTCATGTAGATCAATG
L Q L G V P D N S S T I V T G L L P V H W I C A N Q S T S S Y
Spike
CCAGCCAACGGCTTTTTCTATATTGATGTTGGTAAACACCGTAGTGCCTTTGCACTCCATAGTGGTTATTATGATGCTAACCAGTATTATAT
GGTCGGTTGCCGAAAAAGATATAACTACAACCATTTGTGGCATCACGGAAACGTGAGGTATCACCAATAATACTACGATTGGTCATAATATA
P A N G F F Y I D V G K H R S A F A L H S G Y Y D A N Q Y Y I
Spike
TTATCTCACTAATAAAATACATTTAAATGCTCCTGTCACTCTGAAGATTTGTAAGTTTGGAAACACTTCTTTTGATTTTTTAAGTAATGTTT
AATAGAGTGATTATTTTATGTAAATTTACGAGGACAGTGAGACTTCTAAACATTCAAACCTTTGTGAAGAAAACTAAAAAATTCATTACAAA
Y L T N K I H L N A P V T L K I C K F G N T S F D F L S N V
Spike
CTACTTCTCATGATTGTATAGTTAATTTGTCATTCACAGAACAGTTAGGTGTGCCTTTGGGCATAACTATATCGGGTGAAACTGTACGTTTG
GATGAAGAGTACTAACATATCAATTAAACAGTAAGTGTCTTGTCAATCCACACGGAAACCCGTATTGATATAGCCCACTTTGACATGCAAAC
S T S H D C I V N L S F T E Q L G V P L G I T I S G E T V R L
Spike
CATTTATATAATGCAACTCGTACTTTTTATGTGCCGGCCGCTTATAAACTTACTAAACTTAGTGTTAAATGTTACTTTAGTGAATCCTGTGT
GTAAATATATTACGTTGAGCATGAAAAATACACGGCCGGCGAATATTTGAATGATTTGAATCACAATTTACAATGAAATCACTTAGGACACA
H L Y N A T R T F Y V P A A Y K L T K L S V K C Y F S E S C V
Spike
TTTTAGTGTTGTCAATGCCACCATTACTGTTAATGTCACCACACTTAATGGCCGTATAGTTAACTACACTGTTTGTGATGATTGTAATGGTT
AAAATCACAACAGTTACGGTGGTAATGACAATTACAGTGGTGTGAATTACCGGCATATCAATTGATGTGACAAACACTACTAACATTACCAA
F S V V N A T I T V N V T T L N G R I V N Y T V C D D C N G
Spike
ATACTGATAACATATTTTCTGTTCAACAGGATGGCCGCATTCCTAATGGTTTCCCTTTTAATAATTGGTTTTTGTTAACTAATGGTTCCACA
TATGACTATTGTATAAAAGACAAGTTGTCCTACCGGCGTAAGGATTACCAAAGGGAAAATTATTAACCAAAAACAATTGATTACCAAGGTGT
Y T D N I F S V Q Q D G R I P N G F P F N N W F L L T N G S T
Spike

2703

```
TTAGTGGACGGGGTCTCTAGACTTTATCAACCACTCCGTTTAACTTGTTTATGGCCTGTACCTGGTCTTAAATCTTCAACTGGTTTTGTT
AATCACCTGCCCCAGAGATCTGAAATAGTTGGTGAGGCAAATTGAACAAATACCGGACATGGACCAGAATTTAGAAGTTGACCAAAACAA

L V D G V S R L Y Q P L R L T C L W P V P G L K S S T G F V
                              Spike

TTTTAATGCCACTGGTTCTGATGTTAATTGTAACGGCTATCAACATAATTCTGTTGCTGATGTTATGCGTTACAATCTTAACCTCAGTGCT
AAAATTACGGTGACCAAGACTACAATTAACATTGCCGATAGTTGTATTAAGACAACGACTACAATACGCAATGTTAGAATTGGAGTCACGA

F N A T G S D V N C N G Y Q H N S V A D V M R Y N L N L S A
                              Spike

ATTCTGTGGACAATCTTAAGAGTGGTGTTATAGTTTTTAAAACTTTACAGTACGATGTTTTGTTTTATTGTAGTAATTCTTCTTCAGGTGT
TAAGACACCTGTTAGAATTCTCACCACAATATCAAAAATTTTGAAATGTCATGCTACAAAACAAAATAACATCATTAAGAAGAAGTCCACA

N S V D N L K S G V I V F K T L Q Y D V L F Y C S N S S S G V
                              Spike

CTTGACACCACAATACCTTTTGGCCCTTCCTCTCAACCTTATTACTGTTTTATAAACAGTACTATCAACACTACTCATGTTAGCACTTTTG
GAACTGTGGTGTTATGGAAAACCGGGAAGGAGAGTTGGAATAATGACAAAATATTTGTCATGATAGTTGTGATGAGTACAATCGTGAAAAC

L D T T I P F G P S S Q P Y Y C F I N S T I N T T H V S T F
                              Spike

GGGTATTTTACCACCCACTGTGCGTGAAATTGTTGTTGCTAGAACTGGTCAGTTTTATATTAATGGTTTTAAGTATTTCGATTTGGGTTTC
CCCATAAAATGGTGGGTGACACGCACTTTAACAACAACGATCTTGACCAGTCAAAATATAATTACCAAAATTCATAAAGCTAAACCCAAAG

G I L P P T V R E I V V A R T G Q F Y I N G F K Y F D L G F
                              Spike

TAGAAGCTGTCAATTTTAATGTCACGACTGCTAGTGCCACAGATTTTTGGACGGTTGCATTTGCTACTTTTGTTGATGTTTTGGTTAATGTT
ATCTTCGACAGTTAAAATTACAGTGCTGACGATCACGGTGTCTAAAAACCTGCCAACGTAAACGATGAAAACAACTACAAAACCAATTACAA

I E A V N F N V T T A S A T D F W T V A F A T F V D V L V N V
                              Spike

AGTGCAACTAACATTCAAAACTTACTTTATTGCGATTCTCCATTTGAAAAGTTGCAGTGTGAGCACTTGCAGTTTGGATTGCAAGATGGTTT
TCACGTTGATTGTAAGTTTTGAATGAAATAACGCTAAGAGGTAAACTTTTCAACGTCACACTCGTGAACGTCAAACCTAACGTTCTACCAAA

S A T N I Q N L L Y C D S P F E K L Q C E H L Q F G L Q D G F
                              Spike

TTATTCTGCAAATTTTCTTGATGATAATGTTTTGCCTGAGACTTATGTTGCACTCCCCATTTATTATCAACATACGGACATAAATTTTACTG
AATAAGACGTTTAAAAGAACTACTATTACAAAACGGACTCTGAATACAACGTGAGGGGTAAATAATAGTTGTATGCCTGTATTTAAAATGAC

Y S A N F L D D N V L P E T Y V A L P I Y Y Q H T D I N F T
                              Spike

CAACTGCATCTTTTGGTGGTTCTTGTTATGTTTGTAAACCACGCCAGGTTAATATATCTCTTAATGGTAACACTTCAGTGTGTGTTAGAACA
GTTGACGTAGAAAACCACCAAGAACAATACAAACATTTGGTGCGGTCCAATTATATAGAGAATTACCATTGTGAAGTCACACACAATCTTGT

A T A S F G G S C Y V C K P R Q V N I S L N G N T S V C V R T
                              Spike
```

TCTCATTTTTCAATTAGGTATATTTATAACCGCGTTAAGAGTGGTTCACCAGGTGACTCTTCATGGCATATTTATTTAAAGAGTGGCACTTG
AGAGTAAAAAGTTAATCCATATAAATATTGGCGCAATTCTCACCAAGTGGTCCACTGAGAAGTACCGTATAAATAAATTTCTCACCGTGAAC

S H F S I R Y I Y N R V K S G S P G D S S W H I Y L K S G T C
Spike

TCCATTTTCTTTTTCTAAGTTAAATAATTTTCAAAAGTTTAAGACTATTTGTTTCTCAACCGTCGAAGTGCCTGGTAGTTGTAATTTTCCAC
AGGTAAAAGAAAAAGATTCAATTTATTAAAAGTTTTCAAATTCTGATAAACAAAGAGTTGGCAGCTTCACGGACCATCAACATTAAAAGGTG

P F S F S K L N N F Q K F K T I C F S T V E V P G S C N F P
Spike

TTGAAGCCACCTGGCATTACACTTCTTATACTATTGTTGGTGCTTTGTATGTTACTTGGTCTGAAGGTAATTCCATTACTGGTGTACCTTAT
AACTTCGGTGGACCGTAATGTGAAGAATATGATAACAACCACGAAACATACAATGAACCAGACTTCCATTAAGGTAATGACCACATGGAATA

L E A T W H Y T S Y T I V G A L Y V T W S E G N S I T G V P Y
Spike

CCTGTCTCTGGTATTCGTGAGTTTAGTAATTTAGTTTTAAATAATTGTACCAAATATAATATTTATGATTATGTTGGTACTGGAATTATACG
GGACAGAGACCATAAGCACTCAAATCATTAAATCAAAATTTATTAACATGGTTTATATTATAAATACTAATACAACCATGACCTTAATATGC

P V S G I R E F S N L V L N N C T K Y N I Y D Y V G T G I I R
Spike

TTCTTCAAACCAGTCACTTGCTGGTGGTATTACATATGTTTCTAACTCTGGTAATTTACTTGGTTTTAAAAATGTTTCCACTGGTAACATTT
AAGAAGTTTGGTCAGTGAACGACCACCATAATGTATACAAAGATTGAGACCATTAAATGAACCAAAATTTTTACAAAGGTGACCATTGTAAA

S S N Q S L A G G I T Y V S N S G N L L G F K N V S T G N I
Spike

TTATTGTGACACCATGTAACCAACCAGATCAAGTAGCTGTTTATCAACAAAGCATTATTGGTGCCATGACCGCTGTTAATGAGTCTAGATAT
AATAACACTGTGGTACATTGGTTGGTCTAGTTCATCGACAAATAGTTGTTTCGTAATAACCACGGTACTGGCGACAATTACTCAGATCTATA

F I V T P C N Q P D Q V A V Y Q Q S I I G A M T A V N E S R Y
Spike

GGCTTGCAAAACTTACTACAGTTACCTAACTTTTATTATGTTAGTAATGGTGGTAACAATTGCACTACGGCTGTTATGATTTATTCTAATTT
CCGAACGTTTTGAATGATGTCAATGGATTGAAAATAATACAATCATTACCACCATTGTTAACGTGATGCCGACAATACTAAATAAGATTAAA

G L Q N L L Q L P N F Y Y V S N G G N N C T T A V M I Y S N F
Spike

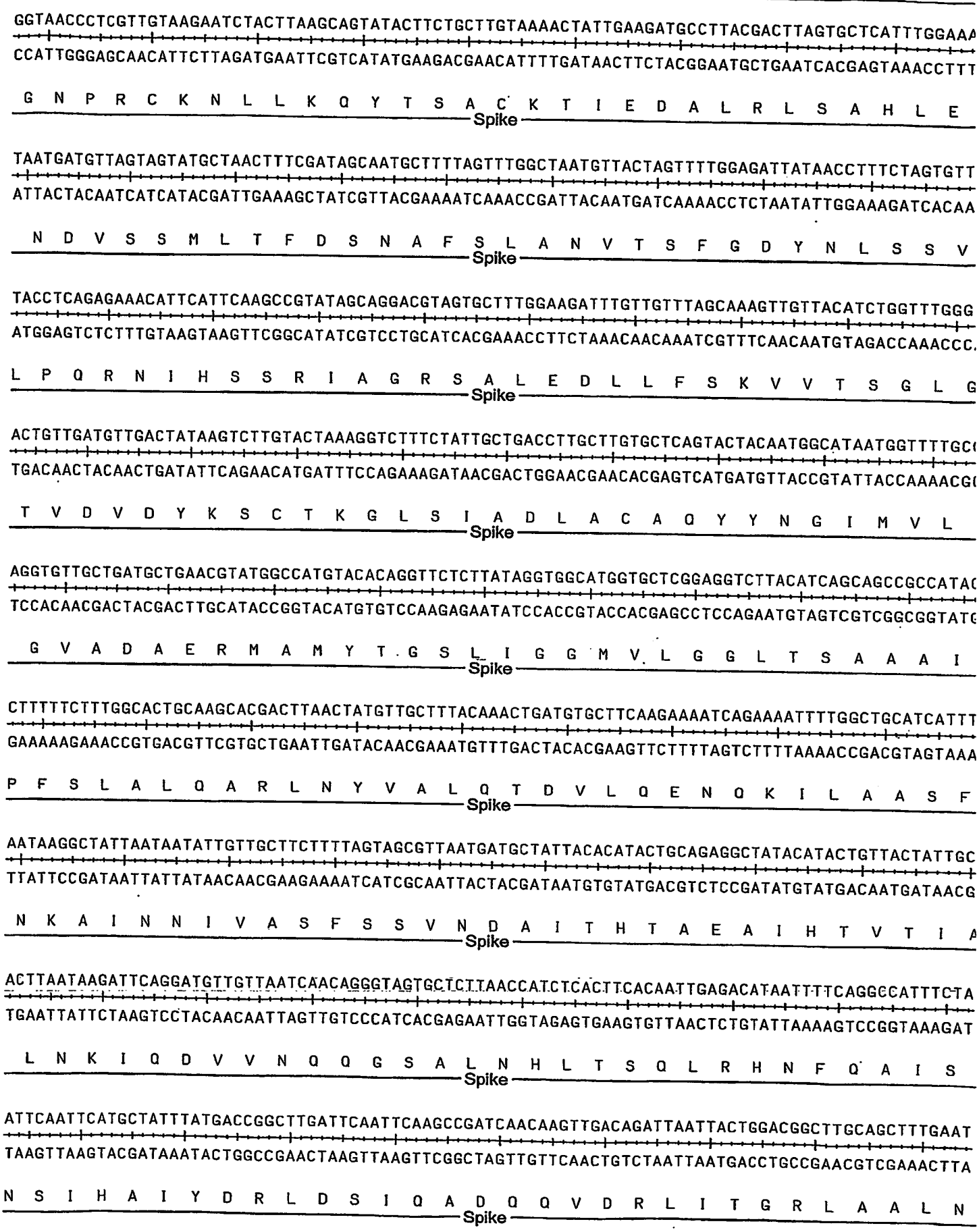
TGGTATTTGTGCTGATGGTTCTTTAATTCCTGTTCGTCCGCGTAATTCTAGTGATAATGGTATTTCAGCCATAATCACTGCTAATTTATCCA
ACCATAAACACGACTACCAAGAAATTAAGGACAAGCAGGCGCATTAAGATCACTATTACCATAAAGTCGGTATTAGTGACGATTAAATAGGT

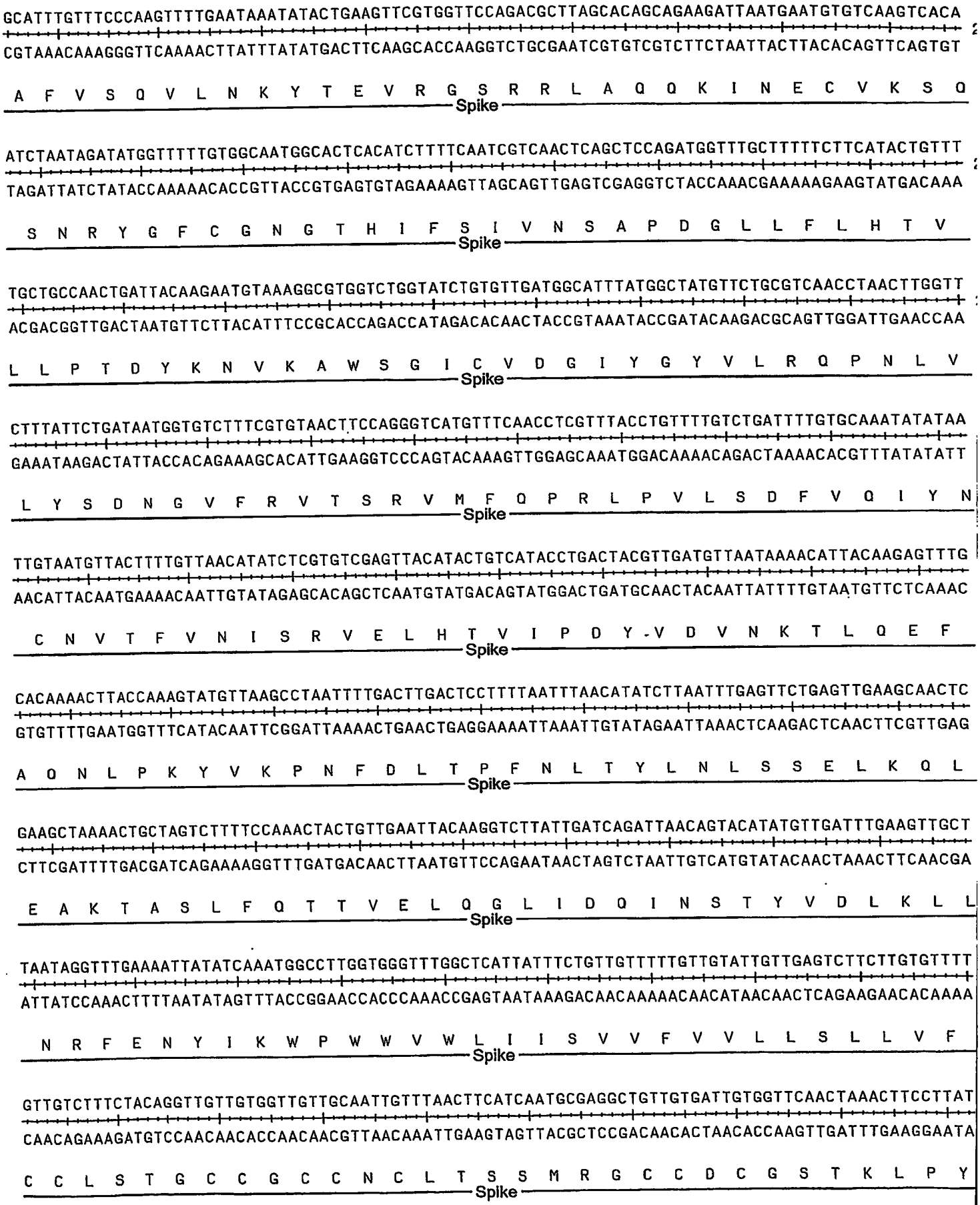
G I C A D G S L I P V R P R N S S D N G I S A I I T A N L S
Spike

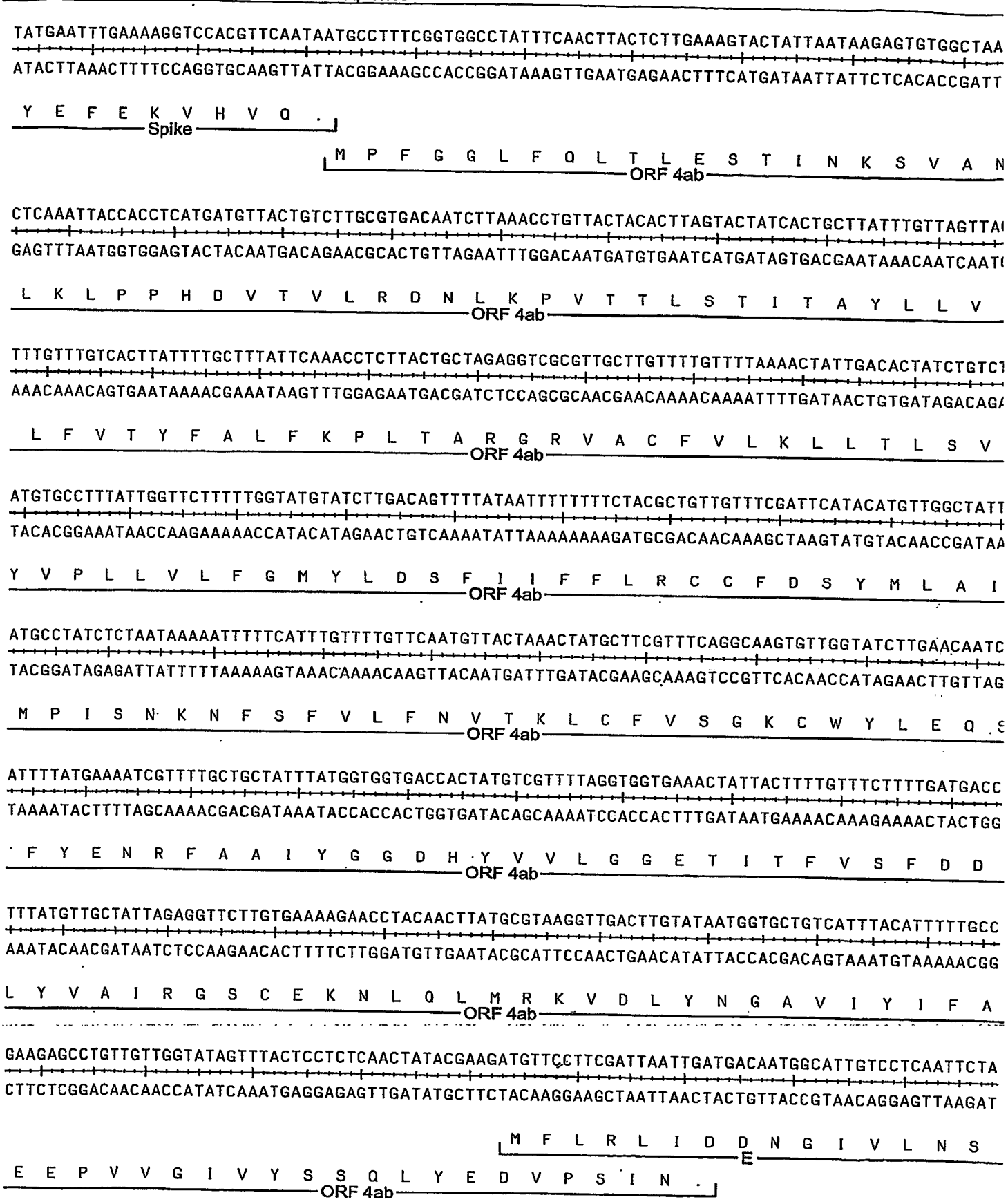
TTCCCTCTAACTGGACTACTTCAGTTCAAGTTGAGTACCTCCAAATTACTAGTACTCCAATAGTTGTTGATTGTGCTACTTATGTGTGTAAT
AAGGGAGATTGACCTGATGAAGTCAAGTTCAACTCATGGAGGTTTAATGATCATGAGGTTATCAACAACTAACACGATGAATACACACATTA

I P S N W T T S V Q V E Y L Q I T S T P I V V D C A T Y V C N
Spike

29107


GGTAACCCTCGTTGTAAGAATCTACTTAAGCAGTATACTTCTGCTTGTAAAACTATTGAAGATGCCTTACGACTTAGTGCTCATTTGGAAA
CCATTGGGAGCAACATTCTTAGATGAATTCGTCATATGAAGACGAACATTTTGATAACTTCTACGGAATGCTGAATCACGAGTAAACCTTT
G N P R C K N L L K Q Y T S A C K T I E D A L R L S A H L E
Spike
TAATGATGTTAGTAGTATGCTAACTTTCGATAGCAATGCTTTTAGTTTGGCTAATGTTACTAGTTTTGGAGATTATAACCTTTCTAGTGTT
ATTACTACAATCATCATACGATTGAAAGCTATCGTTACGAAAATCAAACCGATTACAATGATCAAAACCTCTAATATTGGAAAGATCACAA
N D V S S M L T F D S N A F S L A N V T S F G D Y N L S S V
Spike
TACCTCAGAGAAACATTCATTCAAGCCGTATAGCAGGACGTAGTGCTTTGGAAGATTTGTTGTTTAGCAAAGTTGTTACATCTGGTTTGGG
ATGGAGTCTCTTTGTAAGTAAGTTCGGCATATCGTCCTGCATCACGAAACCTTCTAAACAACAAATCGTTTCAACAATGTAGACCAAACCC
L P Q R N I H S S R I A G R S A L E D L L F S K V V T S G L G
Spike
ACTGTTGATGTTGACTATAAGTCTTGTACTAAAGGTCTTTCTATTGCTGACCTTGCTTGTGCTCAGTACTACAATGGCATAATGGTTTTGC
TGACAACTACAACTGATATTCAGAACATGATTTCCAGAAAGATAACGACTGGAACGAACACGAGTCATGATGTTACCGTATTACCAAAACG
T V D V D Y K S C T K G L S I A D L A C A Q Y Y N G I M V L
Spike
AGGTGTTGCTGATGCTGAACGTATGGCCATGTACACAGGTTCTCTTATAGGTGGCATGGTGCTCGGAGGTCTTACATCAGCAGCCGCCATAC
TCCACAACGACTACGACTTGCATACCGGTACATGTGTCCAAGAGAATATCCACCGTACCACGAGCCTCCAGAATGTAGTCGTCGGCGGTATG
G V A D A E R M A M Y T G S L I G G M V L G G L T S A A A I
Spike
CTTTTTCTTTGGCACTGCAAGCACGACTTAACTATGTTGCTTTACAAACTGATGTGCTTCAAGAAAATCAGAAAATTTTGGCTGCATCATTT
GAAAAAGAAACCGTGACGTTCGTGCTGAATTGATACAACGAAATGTTTGACTACACGAAGTTCTTTTAGTCTTTTAAAACCGACGTAGTAAA
P F S L A L Q A R L N Y V A L Q T D V L Q E N Q K I L A A S F
Spike
AATAAGGCTATTAATAATATTGTTGCTTCTTTTAGTAGCGTTAATGATGCTATTACACATACTGCAGAGGCTATACATACTGTTACTATTGC
TTATTCCGATAATTATTATAACAACGAAGAAAATCATCGCAATTACTACGATAATGTGTATGACGTCTCCGATATGTATGACAATGATAACG
N K A I N N I V A S F S S V N D A I T H T A E A I H T V T I A
Spike
ACTTAATAAGATTCAGGATGTTGTTAATCAACAGGGTAGTGCTCTTAACCATCTCACTTCACAATTGAGACATAATTTTCAGGCCATTTCTA
TGAATTATTCTAAGTCCTACAACAATTAGTTGTCCCATCACGAGAATTGGTAGAGTGAAGTGTTAACTCTGTATTAAAAGTCCGGTAAAGAT
L N K I Q D V V N Q Q G S A L N H L T S Q L R H N F Q A I S
Spike
ATTCAATTCATGCTATTTATGACCGGCTTGATTCAATTCAAGCCGATCAACAAGTTGACAGATTAATTACTGGACGGCTTGCAGCTTTGAAT
TAAGTTAAGTACGATAAATACTGGCCGAACTAAGTTAAGTTCGGCTAGTTGTTCAACTGTCTAATTAATGACCTGCCGAACGTCGAAACTTA
N S I H A I Y D R L D S I Q A D Q Q V D R L I T G R L A A L N
Spike


GCATTTGTTTCCCAAGTTTTGAATAAATATACTGAAGTTCGTGGTTCCAGACGCTTAGCACAGCAGAAGATTAATGAATGTGTCAAGTCACA
CGTAAACAAAGGGTTCAAAACTTATTTATATGACTTCAAGCACCAAGGTCTGCGAATCGTGTCGTCTTCTAATTACTTACACAGTTCAGTGT
A F V S Q V L N K Y T E V R G S R R L A Q Q K I N E C V K S Q
Spike
ATCTAATAGATATGGTTTTTGTGGCAATGGCACTCACATCTTTTCAATCGTCAACTCAGCTCCAGATGGTTTGCTTTTTCTTCATACTGTTT
TAGATTATCTATACCAAAAACACCGTTACCGTGAGTGTAGAAAAGTTAGCAGTTGAGTCGAGGTCTACCAAACGAAAAAGAAGTATGACAAA
S N R Y G F C G N G T H I F S I V N S A P D G L L F L H T V
Spike
TGCTGCCAACTGATTACAAGAATGTAAAGGCGTGGTCTGGTATCTGTGTTGATGGCATTTATGGCTATGTTCTGCGTCAACCTAACTTGGTT
ACGACGGTTGACTAATGTTCTTACATTTCCGCACCAGACCATAGACACAACTACCGTAAATACCGATACAAGACGCAGTTGGATTGAACCAA
L L P T D Y K N V K A W S G I C V D G I Y G Y V L R Q P N L V
Spike
CTTTATTCTGATAATGGTGTCTTTCGTGTAACTTCCAGGGTCATGTTTCAACCTCGTTTACCTGTTTTGTCTGATTTTGTGCAAATATATAA
GAAATAAGACTATTACCACAGAAAGCACATTGAAGGTCCCAGTACAAAGTTGGAGCAAATGGACAAAACAGACTAAAACACGTTTATATATT
L Y S D N G V F R V T S R V M F Q P R L P V L S D F V Q I Y N
Spike
TTGTAATGTTACTTTTGTTAACATATCTCGTGTCGAGTTACATACTGTCATACCTGACTACGTTGATGTTAATAAAACATTACAAGAGTTTG
AACATTACAATGAAAACAATTGTATAGAGCACAGCTCAATGTATGACAGTATGGACTGATGCAACTACAATTATTTTGTAATGTTCTCAAAC
C N V T F V N I S R V E L H T V I P D Y-V D V N K T L Q E F
Spike
CACAAAACTTACCAAAGTATGTTAAGCCTAATTTTGACTTGACTCCTTTTAATTTAACATATCTTAATTTGAGTTCTGAGTTGAAGCAACTC
GTGTTTTGAATGGTTTCATACAATTCGGATTAAAACTGAACTGAGGAAAATTAAATTGTATAGAATTAAACTCAAGACTCAACTTCGTTGAG
A Q N L P K Y V K P N F D L T P F N L T Y L N L S S E L K Q L
Spike
GAAGCTAAAACTGCTAGTCTTTTCCAAACTACTGTTGAATTACAAGGTCTTATTGATCAGATTAACAGTACATATGTTGATTTGAAGTTGCT
CTTCGATTTTGACGATCAGAAAAGGTTTGATGACAACTTAATGTTCCAGAATAACTAGTCTAATTGTCATGTATACAACTAAACTTCAACGA
E A K T A S L F Q T T V E L Q G L I D Q I N S T Y V D L K L L
Spike
TAATAGGTTTGAAAATTATATCAAATGGCCTTGGTGGGTTTGGCTCATTATTTCTGTTGTTTTTGTTGTATTGTTGAGTCTTCTTGTGTTTT
ATTATCCAAACTTTTAATATAGTTTACCGGAACCACCCAAACCGAGTAATAAAGACAACAAAAACAACATAACAACTCAGAAGAACACAAAA
N R F E N Y I K W P W W V W L I I S V V F V V L L S L L V F
Spike
GTTGTCTTTCTACAGGTTGTTGTGGTTGTTGCAATTGTTTAACTTCATCAATGCGAGGCTGTTGTGATTGTGGTTCAACTAAACTTCCTTAT
CAACAGAAAGATGTCCAACAACACCAACAACGTTAACAAATTGAAGTAGTTACGCTCCGACAACACTAACACCAAGTTGATTTGAAGGAATA
C C L S T G C C G C C N C L T S S M R G C C D C G S T K L P Y
Spike


TATGAATTTGAAAAGGTCCACGTTCAATAATGCCTTTCGGTGGCCTATTTCAACTTACTCTTGAAAGTACTATTAATAAGAGTGTGGCTAA
ATACTTAAACTTTTCCAGGTGCAAGTTATTACGGAAAGCCACCGGATAAAGTTGAATGAGAACTTTCATGATAATTATTCTCACACCGATT
Y E F E K V H V Q .
Spike
M P F G G L F Q L T L E S T I N K S V A N
ORF 4ab
CTCAAATTACCACCTCATGATGTTACTGTCTTGCGTGACAATCTTAAACCTGTTACTACACTTAGTACTATCACTGCTTATTTGTTAGTTA
GAGTTTAATGGTGGAGTACTACAATGACAGAACGCACTGTTAGAATTTGGACAATGATGTGAATCATGATAGTGACGAATAAACAATCAAT
L K L P P H D V T V L R D N L K P V T T L S T I T A Y L L V
ORF 4ab
TTTGTTTGTCACTTATTTTGCTTTATTCAAACCTCTTACTGCTAGAGGTCGCGTTGCTTGTTTTGTTTTAAAACTATTGACACTATCTGTC
AAACAAACAGTGAATAAAACGAAATAAGTTTGGAGAATGACGATCTCCAGCGCAACGAACAAAACAAAATTTTGATAACTGTGATAGACAG
L F V T Y F A L F K P L T A R G R V A C F V L K L L T L S V
ORF 4ab
ATGTGCCTTTATTGGTTCTTTTTGGTATGTATCTTGACAGTTTTATAATTTTTTTTCTACGCTGTTGTTTCGATTCATACATGTTGGCTATT
TACACGGAAATAACCAAGAAAAACCATACATAGAACTGTCAAAATATTAAAAAAAAGATGCGACAACAAAGCTAAGTATGTACAACCGATAA
Y V P L L V L F G M Y L D S F I I F F L R C C F D S Y M L A I
ORF 4ab
ATGCCTATCTCTAATAAAAATTTTTCATTTGTTTTGTTCAATGTTACTAAACTATGCTTCGTTTCAGGCAAGTGTTGGTATCTTGAACAATC
TACGGATAGAGATTATTTTTAAAAAGTAAACAAAACAAGTTACAATGATTTGATACGAAGCAAAGTCCGTTCACAACCATAGAACTTGTTAG
M P I S N K N F S F V L F N V T K L C F V S G K C W Y L E Q .
ORF 4ab
ATTTTATGAAAATCGTTTTGCTGCTATTTATGGTGGTGACCACTATGTCGTTTTAGGTGGTGAAACTATTACTTTTGTTTCTTTTGATGACC
TAAAATACTTTTAGCAAAACGACGATAAATACCACCACTGGTGATACAGCAAAATCCACCACTTTGATAATGAAAACAAAGAAAACTACTGG
F Y E N R F A A I Y G G D H Y V V L G G E T I T F V S F D D
ORF 4ab
TTTATGTTGCTATTAGAGGTTCTTGTGAAAAGAACCTACAACTTATGCGTAAGGTTGACTTGTATAATGGTGCTGTCATTTACATTTTTGCC
AAATACAACGATAATCTCCAAGAACACTTTTCTTGGATGTTGAATACGCATTCCAACTGAACATATTACCACGACAGTAAATGTAAAAACGG
L Y V A I R G S C E K N L Q L M R K V D L Y N G A V I Y I F A
ORF 4ab
GAAGAGCCTGTTGTTGGTATAGTTTACTCCTCTCAACTATACGAAGATGTTCCTTCGATTAATTGATGACAATGGCATTGTCCTCAATTCTA
CTTCTCGGACAACAACCATATCAAATGAGGAGAGTTGATATGCTTCTACAAGGAAGCTAATTAACTACTGTTACCGTAACAGGAGTTAAGAT
M F L R L I D D N G I V L N S
E
E E P V V G I V Y S S Q L Y E D V P S I N .
ORF 4ab

TTTTATGGCTCCTTGTTATGATATTTTTCTTTGTGTTGGCAATGACCTTTATTAAACTGATTCAATTGTGTTTTACTTGTCATTATTTTTTT 2
AAAATACCGAGGAACAATACTATAAAAAGAAACACAACCGTTACTGGAAATAATTTGACTAAGTTAACACAAAATGAACAGTAATAAAAAAA

I L W L L V M I F F F V L A M T F I K L I Q L C F T C H Y F F
E

AGTAGGACATTATATCAACCAGTTTATAAAATTTTTCTTGCTTACCAAGATTATATGCAAATAGCACCTGTTCCAGCTGAAGTACTAAATGT 2
TCATCCTGTAATATAGTTGGTCAAATATTTTAAAAAGAACGAATGGTTCTAATATACGTTTATCGTGGACAAGGTCGACTTCATGATTTACA

S R T L Y Q P V Y K I F L A Y Q D Y M Q I A P V P A E V L N V
E

CTAAACTAAACGATGTCTAATAGTAGTGTGCCTCTTTCAGAGGTTTATGTCCATTTACGTAACTGGAACTTTAGTTGGAATTTAATTCTAAC 2
GATTTGATTTGCTACAGATTATCATCACACGGAGAAAGTCTCCAAATACAGGTAAATGCATTGACCTTGAAATCAACCTTAAATTAAGATTG

E M S N S S V P L S E V Y V H L R N W N F S W N L I L T
M

AGTTTTTATAGTTGTGTTGCAGTATGGGCATTATAAGTATAGCAGACTTCTTTATGGTTTAAAGATGTCTGTTTTATGGTGTTTATGGCCAC 2
TCAAAAATATCAACACAACGTCATACCCGTAATATTCATATCGTCTGAAGAAATACCAAATTTCTACAGACAAAATACCACAAATACCGGTG

V F I V V L Q Y G H Y K Y S R L L Y G L K M S V L W C L W P
M

TTGTTCTAGCTTTGTCTATTTTTGACTGTTTTGTCAATTTTAATGTGGACTGGGTCTTTTTTGGTTTTAGTATTCTTATGTCTATTATTACA 2
AACAAGATCGAAACAGATAAAAACTGACAAAACAGTTAAAATTACACCTGACCCAGAAAAAACCAAAATCATAAGAATACAGATAATAATGT

L V L A L S I F D C F V N F N V D W V F F G F S I L M S I I T
M

CTTTGTTTATGGGTTATGTATTTTGTTAATAGTTTCAGACTTTGGCGCCGTGTTAAAACTTTTTGGGCTTTTAATCCTGAAACTAATGCAAT 2
GAAACAAATACCCAATACATAAAACAATTATCAAAGTCTGAAACCGCGGCACAATTTTGAAAAACCCGAAAATTAGGACTTTGATTACGTTA

L C L W V M Y F V N S F R L W R R V K T F W A F N P E T N A I
M

CATCTCTCTCCAGGTTTATGGACATAATTATTACTTACCGGTGATGGCTGCACCTACAGGTGTTACATTAACACTTCTTAGTGGTGTACTTC
GTAGAGAGAGGTCCAAATACCTGTATTAATAATGAATGGCCACTACCGACGTGGATGTCCACAATGTAATTGTGAAGAATCACCACATGAAG

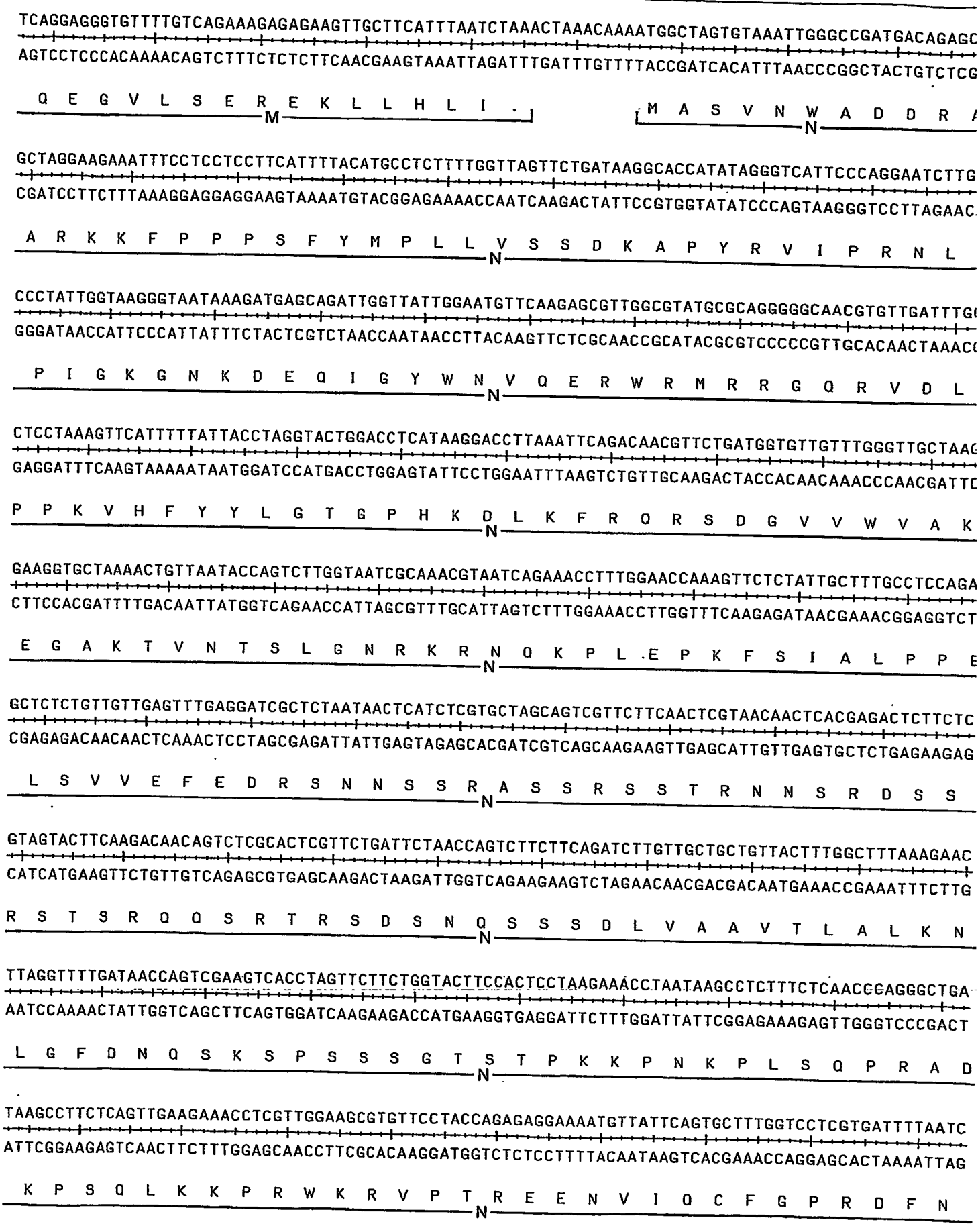
I S L Q V Y G H N Y Y L P V M A A P T G V T L T L L S G V L
M

TTGTTGATGGCCATAAGATTGCTACTCGTGTTCAAGTGGGTCAGTTGCCTAAATATGTAATAGTTGCTACACCTAGTACCACAATTGTTTGT
AACAACTACCGGTATTCTAACGATGAGCACAAGTTCACCCAGTCAACGGATTTATACATTATCAACGATGTGGATCATGGTGTTAACAAACA

L V D G H K I A T R V Q V G Q L P K Y V I V A T P S T T I V C
M

GACCGTGTTGGTCGCTCTGTTAATGAAACAAGCCAGACTGGTTGGGCATTCTACGTCCGTGCTAAACATGGTGATTTTTCTGGTGTTGCCTC
CTGGCACAACCAGCGAGACAATTACTTTGTTCGGTCTGACCAACCCGTAAGATGCAGGCACGATTTGTACCACTAAAAAGACCACAACGGAG

D R V G R S V N E T S Q T G W A F Y V R A K H G D F S G V A S
M


TCAGGAGGGTGTTTTGTCAGAAAGAGAGAAGTTGCTTCATTTAATCTAAACTAAACAAAATGGCTAGTGTAAATTGGGCCGATGACAGAGC
AGTCCTCCCACAAAACAGTCTTTCTCTCTTCAACGAAGTAAATTAGATTTGATTTGTTTTACCGATCACATTTAACCCGGCTACTGTCTCG
Q E G V L S E R E K L L H L I .
M
M A S V N W A D D R A
N
GCTAGGAAGAAATTTCCTCCTCCTTCATTTTACATGCCTCTTTTGGTTAGTTCTGATAAGGCACCATATAGGGTCATTCCCAGGAATCTTG
CGATCCTTCTTTAAAGGAGGAGGAAGTAAAATGTACGGAGAAAACCAATCAAGACTATTCCGTGGTATATCCCAGTAAGGGTCCTTAGAAC
A R K K F P P P S F Y M P L L V S S D K A P Y R V I P R N L
N
CCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCGCAGGGGGCAACGTGTTGATTTG
GGGATAACCATTCCCATTATTTCTACTCGTCTAACCAATAACCTTACAAGTTCTCGCAACCGCATACGCGTCCCCCGTTGCACAACTAAAC
P I G K G N K D E Q I G Y W N V Q E R W R M R R G Q R V D L
N
CTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTTGTTTGGGTTGCTAAG
GAGGATTTCAAGTAAAAATAATGGATCCATGACCTGGAGTATTCCTGGAATTTAAGTCTGTTGCAAGACTACCACAACAAACCCAACGATTC
P P K V H F Y Y L G T G P H K D L K F R Q R S D G V V W V A K
N
GAAGGTGCTAAAACTGTTAATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCTCCAGA
CTTCCACGATTTTGACAATTATGGTCAGAACCATTAGCGTTTGCATTAGTCTTTGGAAACCTTGGTTTCAAGAGATAACGAAACGGAGGTCT
E G A K T V N T S L G N R K R N Q K P L E P K F S I A L P P E
N
GCTCTCTGTTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGCTAGCAGTCGTTCTTCAACTCGTAACAACTCACGAGACTCTTCTC
CGAGAGACAACAACTCAAACTCCTAGCGAGATTATTGAGTAGAGCACGATCGTCAGCAAGAAGTTGAGCATTGTTGAGTGCTCTGAGAAGAG
L S V V E F E D R S N N S S R A S S R S S T R N N S R D S S
N
GTAGTACTTCAAGACAACAGTCTCGCACTCGTTCTGATTCTAACCAGTCTTCTTCAGATCTTGTTGCTGCTGTTACTTTGGCTTTAAAGAAC
CATCATGAAGTTCTGTTGTCAGAGCGTGAGCAAGACTAAGATTGGTCAGAAGAAGTCTAGAACAACGACGACAATGAAACCGAAATTTCTTG
R S T S R Q Q S R T R S D S N Q S S S D L V A A V T L A L K N
N
TTAGGTTTTGATAACCAGTCGAAGTCACCTAGTTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTCTTTCTCAACCGAGGGCTGA
AATCCAAAACTATTGGTCAGCTTCAGTGGATCAAGAAGACCATGAAGGTGAGGATTCTTTGGATTATTCGGAGAAAGAGTTGGGTCCCGACT
L G F D N Q S K S P S S S G T S T P K K P N K P L S Q P R A D
N
TAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTTCCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCCTCGTGATTTTAATC
ATTCGGAAGAGTCAACTTCTTTGGAGCAACCTTCGCACAAGGATGGTCTCTCCTTTTACAATAAGTCACGAAACCAGGAGCACTAAAATTAG
K P S Q L K K P R W K R V P T R E E N V I Q C F G P R D F N
N

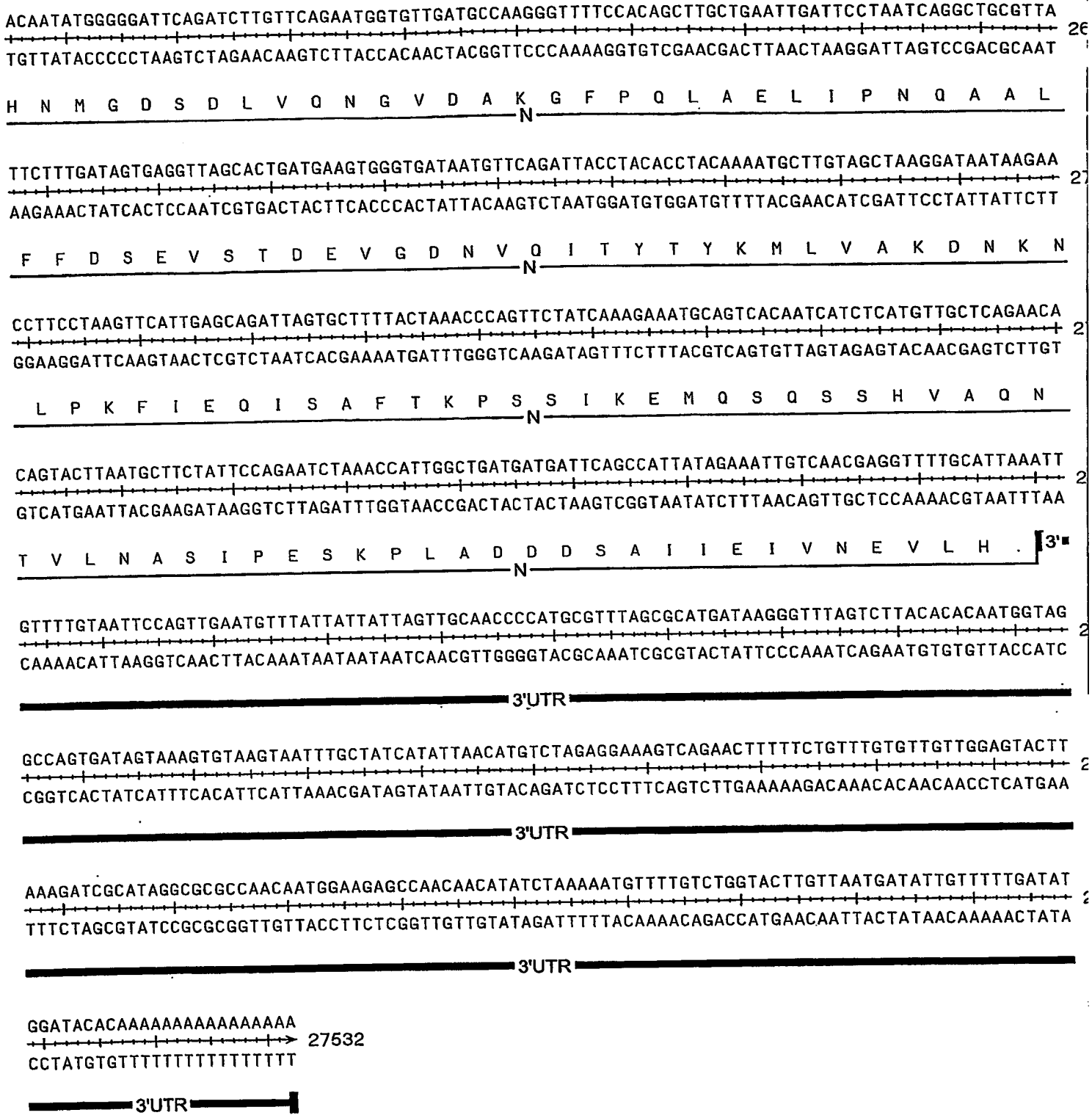

ACAATATGGGGGATTCAGATCTTGTTCAGAATGGTGTTGATGCCAAGGGTTTTCCACAGCTTGCTGAATTGATTCCTAATCAGGCTGCGTTA
TGTTATACCCCCTAAGTCTAGAACAAGTCTTACCACAACTACGGTTCCCAAAAGGTGTCGAACGACTTAACTAAGGATTAGTCCGACGCAAT
H N M G D S D L V Q N G V D A K G F P Q L A E L I P N Q A A L
N
TTCTTTGATAGTGAGGTTAGCACTGATGAAGTGGGTGATAATGTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAA
AAGAAACTATCACTCCAATCGTGACTACTTCACCCACTATTACAAGTCTAATGGATGTGGATGTTTTACGAACATCGATTCCTATTATTCTT
F F D S E V S T D E V G D N V Q I T Y T Y K M L V A K D N K N
N
CCTTCCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGTTCTATCAAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACA
GGAAGGATTCAAGTAACTCGTCTAATCACGAAAATGATTTGGGTCAAGATAGTTTCTTTACGTCAGTGTTAGTAGAGTACAACGAGTCTTGT
L P K F I E Q I S A F T K P S S I K E M Q S Q S S H V A Q N
N
CAGTACTTAATGCTTCTATTCCAGAATCTAAACCATTGGCTGATGATGATTCAGCCATTATAGAAATTGTCAACGAGGTTTTGCATTAAATT
GTCATGAATTACGAAGATAAGGTCTTAGATTTGGTAACCGACTACTACTAAGTCGGTAATATCTTTAACAGTTGCTCCAAAACGTAATTTAA
T V L N A S I P E S K P L A D D D S A I I E I V N E V L H .
N
3'
GTTTTGTAATTCCAGTTGAATGTTTATTATTATTAGTTGCAACCCCATGCGTTTAGCGCATGATAAGGGTTTAGTCTTACACACAATGGTAG
CAAAACATTAAGGTCAACTTACAAATAATAATAATCAACGTTGGGGTACGCAAATCGCGTACTATTCCCAAATCAGAATGTGTGTTACCATC
3'UTR
GCCAGTGATAGTAAAGTGTAAGTAATTTGCTATCATATTAACATGTCTAGAGGAAAGTCAGAACTTTTTCTGTTTGTGTTGTTGGAGTACTT
CGGTCACTATCATTTCACATTCATTAAACGATAGTATAATTGTACAGATCTCCTTTCAGTCTTGAAAAAGACAAACACAACAACCTCATGAA
3'UTR
AAAGATCGCATAGGCGCGCCAACAATGGAAGAGCCAACAACATATCTAAAAATGTTTTGTCTGGTACTTGTTAATGATATTGTTTTTGATAT
TTTCTAGCGTATCCGCGCGGTTGTTACCTTCTCGGTTGTTGTATAGATTTTTACAAAACAGACCATGAACAATTACTATAACAAAAACTATA
3'UTR
GGATACACAAAAAAAAAAAAAAA
27532
CCTATGTGTTTTTTTTTTTTTTT
3'UTR

Figure 2

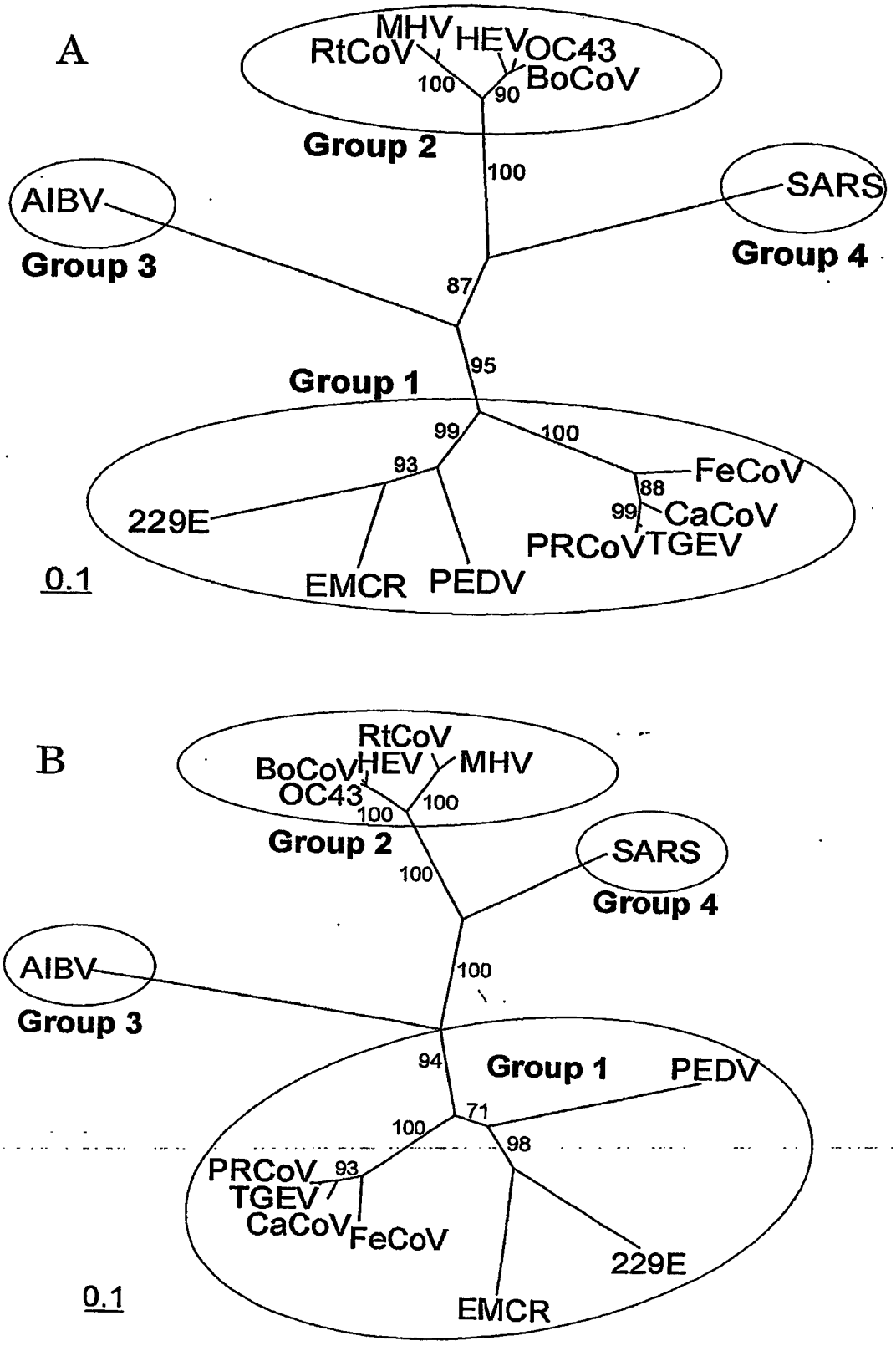

A
MHV
RtCoV
HEV
OC43
BoCoV
100
90
Group 2
100
AIBV
Group 3
SARS
Group 4
87
95
Group 1
99
100
93
FeCoV
88
229E
99
CaCoV
PRCoV
TGEV
0.1
EMCR
PEDV
B
RtCoV
BoCoV
HEV
MHV
OC43
100
100
Group 2
SARS
100
Group 4
AIBV
Group 3
100
94
Group 1
PEDV
71
100
98
PRCoV
93
TGEV
CaCoV
FeCoV
229E
0.1
EMCR

Figure 3a: Putative Orf 1a Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	OC43	BoCoV	MHV	AIPV	SARS
EMCR	1.000	0,566	0,491	0,371	0,211	0,213	0,215	0,185	0,194
229E	---	1.000	0,475	0,379	0,211	0,211	0,209	0,191	0,194
PEDV	---	---	1.000	0,366	0,208	0,208	0,204	0,183	0,192
TGEV	---	---	---	1.000	0,204	0,204	0,199	0,182	0,186
OC43	---	---	---	---	1.000	0,964	0,656	0,182	0,255
BoCoV	---	---	---	---	---	1.000	0,659	0,186	0,255
MHV	---	---	---	---	---	---	1.000	0,178	0,253
AIPV	---	---	---	---	---	---	---	1.000	0,178
SARS	---	---	---	---	---	---	---	---	1.000

Figure 3b: Putative Orf 1b Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	BoCoV	OC43	MHV	AIPV	SARS
EMCR	1.000	0,815	0,778	0,711	0,504	0,517	0,523	0,516	0,550
229E	—	1.000	0,765	0,720	0,504	0,520	0,515	0,515	0,546
PEDV	—	—	1.000	0,728	0,522	0,538	0,531	0,523	0,552
TGEV	—	—	—	1.000	0,517	0,533	0,529	0,520	0,551
BoCoV	—	—	—	—	1.000	0,953	0,832	0,519	0,591
OC43	—	—	—	—	—	1.000	0,857	0,534	0,611
MHV	—	—	—	—	—	—	1.000	0,532	0,606
AIPV	—	—	—	—	—	—	—	1.000	0,541
SARS	—	—	—	—	—	—	—	—	1.000

Figure 3c: Putative Orf 1ab Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	OV43	BoCoV	MHV	AIBV	SARS
EMCR	1.000	0,666	0,605	0,503	0,314	0,310	0,316	0,314	0,326
229E	—	1.000	0,592	0,510	0,320	0,314	0,316	0,311	0,326
PEDV	—	—	1.000	0,509	0,326	0,320	0,323	0,313	0,328
TGEV	—	—	—	1.000	0,317	0,311	0,313	0,311	0,320
OV43	—	—	—	—	1.000	0,961	0,734	0,312	0,408
BoCoV	—	—	—	—	—	1.000	0,725	0,307	0,400
MHV	—	—	—	—	—	—	1.000	0,309	0,404
AIBV	—	—	—	—	—	—	—	1.000	0,312
SARS	—	—	—	—	—	—	—	—	1.000

Figure 3d: Putative Spike protein Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	Por R	OC43	BoCoV	MHV	Rat C	PHEV	AIBV	SARS
EMCR	1.000	0,547	0,442	0,387	0,386	0,394	0,393	0,205	0,206	0,196	0,198	0,203	0,202	0,167
229E	---	1.000	0,412	0,383	0,381	0,383	0,445	0,178	0,185	0,178	0,174	0,179	0,230	0,176
PEDV	---	---	1.000	0,412	0,412	0,415	0,403	0,193	0,188	0,189	0,189	0,194	0,198	0,179
TGEV	---	---	---	1.000	0,787	0,802	0,812	0,196	0,195	0,199	0,202	0,197	0,188	0,171
CaCoV	---	---	---	---	1.000	0,911	0,743	0,196	0,195	0,202	0,202	0,194	0,184	0,167
FeCoV	---	---	---	---	---	1.000	0,758	0,195	0,193	0,197	0,200	0,196	0,188	0,167

Por R	---	---	---	---	---	---	1.000	0,183	0,185	0,189	0,191	0,186	0,212	0,164
OC43	---	---	---	---	---	---	---	1.000	0,911	0,637	0,637	0,803	0,171	0,254
BoCoV	---	---	---	---	---	---	---	---	1.000	0,642	0,643	0,817	0,173	0,255
MHV	---	---	---	---	---	---	---	---	---	1.000	0,909	0,643	0,183	0,249
Rat C	---	---	---	---	---	---	---	---	---	---	1.000	0,646	0,184	0,255
PHEV	---	---	---	---	---	---	---	---	---	---	---	1.000	0,177	0,252
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0,164
SARS	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3e: Putative Orf E Amino acid identity matrix

Seq->	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	Por R	OC43	BoCoV	PHEV	MHV	Rat C	AIBV	SARS
EMCR	1.000	0,467	0,415	0,292	0,304	0,256	0,304	0,154	0,154	0,154	0,181	0,181	0,092	0,179
229E	---	1.000	0,532	0,243	0,231	0,243	0,231	0,214	0,214	0,214	0,227	0,227	0,120	0,230
PEDV	---	---	1.000	0,280	0,268	0,256	0,280	0,214	0,214	0,214	0,227	0,227	0,092	0,179
TGEV	---	---	---	1.000	0,914	0,743	0,963	0,172	0,172	0,172	0,172	0,172	0,145	0,228
CaCoV	---	---	---	---	1.000	0,804	0,926	0,172	0,172	0,172	0,193	0,193	0,127	0,216
FeCoV	---	---	---	---	---	1.000	0,756	0,183	0,183	0,183	0,182	0,182	0,127	0,204
Por R	---	---	---	---	---	---	1.000	0,183	0,183	0,183	0,172	0,172	0,136	0,216
OC43	---	---	---	---	---	---	---	1.000	0,976	0,988	0,644	0,655	0,119	0,176
BoCoV	---	---	---	---	---	---	---	---	1.000	0,988	0,622	0,633	0,110	0,176
PHEV	---	---	---	---	---	---	---	---	---	1.000	0,633	0,644	0,119	0,176
MHV	---	---	---	---	---	---	---	---	---	---	1.000	0,977	0,128	0,191
Rat C	---	---	---	---	---	---	---	---	---	---	---	1.000	0,128	0,202
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0,137
SARS	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3f: Matrix amino acid identity

Seq->	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	PRCoV	OC43	PHEV	BoCoV	MHV	RatSA	AIBV	SARS
EMCR	1.000	0,615	0,650	0,441	0,429	0,400	0,437	0,317	0,317	0,317	0,303	0,303	0,239	0,286
229E	---	1.000	0,557	0,380	0,372	0,344	0,384	0,317	0,309	0,309	0,299	0,316	0,269	0,277
PEDV	---	---	1.000	0,460	0,452	0,386	0,460	0,351	0,360	0,364	0,358	0,363	0,234	0,303
TGEV	---	---	---	1.000	0,878	0,758	0,958	0,330	0,334	0,346	0,335	0,332	0,208	0,254
CaCoV	---	---	---	---	1.000	0,757	0,878	0,311	0,315	0,326	0,319	0,304	0,192	0,243
FeCoV	---	---	---	---	---	1.000	0,772	0,268	0,279	0,286	0,272	0,265	0,174	0,234
PRCoV	---	---	---	---	---	---	1.000	0,330	0,334	0,346	0,335	0,332	0,215	0,262
OC43	---	---	---	---	---	---	---	1.000	0,934	0,947	0,848	0,818	0,270	0,386
PHEV	---	---	---	---	---	---	---	---	1.000	0,943	0,848	0,818	0,270	0,400
BoCoV	---	---	---	---	---	---	---	---	---	1.000	0,870	0,839	0,278	0,391
MHV	---	---	---	---	---	---	---	---	---	---	1.000	0,938	0,271	0,382
RatSA	---	---	---	---	---	---	---	---	---	---	---	1.000	0,275	0,369
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0,262
SARS	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3g Nucleoprotein amino acid identity

Seq->	EMCR	229E	PEDV	TGEV	FeCoV	PRCoV	CaCoV	RSDAC	MHV	PHEV	OC43	BoCoV	SARS	AIBV
EMCR	1.000	0,447	0,363	0,331	0,319	0,329	0,339	0,188	0,189	0,179	0,183	0,183	0,214	0,179
229E	---	1.000	0,345	0,333	0,304	0,326	0,330	0,194	0,202	0,192	0,194	0,192	0,203	0,181
PEDV	---	---	1.000	0,279	0,248	0,275	0,275	0,165	0,170	0,164	0,164	0,162	0,182	0,180
TGEV	---	---	---	1.000										

FeCoV	---	---	---	---	1.000	0,756	0,763	0,212	0,213	0,192	0,195	0,197	0,217	0,185
PRCoV	---	---	---	---	---	1.000	0,879	0,220	0,226	0,205	0,207	0,205	0,231	0,192
CaCoV	---	---	---	---	---	---	1.000	0,216	0,220	0,199	0,201	0,199	0,217	0,199
RSDAC	---	---	---	---	---	---	---	1.000	0,894	0,693	0,697	0,697	0,293	0,208
MHV	---	---	---	---	---	---	---	---	1.000	0,680	0,684	0,682	0,290	0,216
PHEV	---	---	---	---	---	---	---	---	---	1.000	0,948	0,953	0,270	0,199
OC43	---	---	---	---	---	---	---	---	---	---	1.000	0,973	0,272	0,203
BoCoV	---	---	---	---	---	---	---	---	---	---	---	1.000	0,274	0,203
SARS	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0,206
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3h Matrix nucleotide identity

Seq->	SARS	EMCR	229E	PEDV	TGEV	CaCoV	FeCoV	PRCoV	OC43	PHEV	BoCoV	MHV	RatSA	AIBV
SARS	1.000	0.286	0.281	0.303	0.254	0.243	0.258	0.262	0.386	0.400	0.391	0.382	0.369	0.262
EMCR	---	1.000	0.615	0.650	0.441	0.429	0.441	0.437	0.317	0.317	0.317	0.303	0.303	0.239
229E	---	---	1.000	0.557	0.380	0.365	0.376	0.380	0.321	0.313	0.313	0.303	0.320	0.269
PEDV	---	---	---	1.000	0.460	0.452	0.425	0.460	0.351	0.360	0.364	0.358	0.363	0.234
TGEV	---	---	---	---	1.000	0.878	0.836	0.958	0.330	0.334	0.346	0.335	0.332	0.208
CaCoV	---	---	---	---	---	1.000	0.835	0.878	0.311	0.315	0.326	0.319	0.304	0.192
FeCoV	---	---	---	---	---	---	1.000	0.851	0.296	0.307	0.315	0.300	0.292	0.192
PRCoV	---	---	---	---	---	---	---	1.000	0.330	0.334	0.346	0.335	0.332	0.215
OC43	---	---	---	---	---	---	---	---	1.000	0.934	0.947	0.848	0.818	0.270
PHEV	---	---	---	---	---	---	---	---	---	1.000	0.943	0.848	0.818	0.270
BoCoV	---	---	---	---	---	---	---	---	---	---	1.000	0.870	0.839	0.278
MHV	---	---	---	---	---	---	---	---	---	---	---	1.000	0.938	0.271
RatSA	---	---	---	---	---	---	---	---	---	---	---	---	1.000	0.275
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	---	1.000

Figure 3i Matrix nucleotide identity

Seq->	EMCR	229E	PEDV	TGEV	FeCoV	PRCoV	CaCoV	RSDAC	MHV	PHEV	OC43	BoCoV	SARS	AIBV
EMCR	1,000	0,447	0,358	0,336	0,326	0,334	0,344	0,188	0,189	0,179	0,183	0,183	0,210	0,173
229E	---	1,000	0,336	0,335	0,304	0,328	0,333	0,196	0,204	0,187	0,190	0,188	0,199	0,173
PEDV	---	---	1,000	0,277	0,244	0,272	0,270	0,163	0,168	0,160	0,160	0,158	0,184	0,178
TGEV	---	---	---	1,000	0,761	0,963	0,897	0,220	0,223	0,200	0,202	0,200	0,232	0,192
FeCoV	---	---	---	---	1,000	0,756	0,763	0,209	0,212	0,185	0,187	0,189	0,218	0,185
PRCoV	---	---	---	---	---	1,000	0,879	0,220	0,228	0,202	0,204	0,202	0,230	0,192
CaCoV	---	---	---	---	---	---	1,000	0,215	0,221	0,196	0,198	0,196	0,216	0,196
RSDAC	---	---	---	---	---	---	---	1,000	0,894	0,693	0,697	0,697	0,285	0,200
MHV	---	---	---	---	---	---	---	---	1,000	0,680	0,684	0,682	0,282	0,208
PHEV	---	---	---	---	---	---	---	---	---	1,000	0,948	0,953	0,261	0,195
OC43	---	---	---	---	---	---	---	---	---	---	1,000	0,973	0,261	0,197
BoCoV	---	---	---	---	---	---	---	---	---	---	---	1,000	0,266	0,197
SARS	---	---	---	---	---	---	---	---	---	---	---	---	1,000	0,211
AIBV	---	---	---	---	---	---	---	---	---	---	---	---	---	1,000

Figure 4 Alignments

a. 5′ untranslated region (Genomic sequence) aligned with human coronavi:
229E

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         5         15         25         35         45         55
EMCR5'UTR           ---------- ---------- ---AGATAGA GAATTTTCTT ATTTAGACTT TGTGTCTACT
229E5'UTR           ACTTAAGTAC CTTATCTATC TACAGATAGA AAAGTTGCTT -TTTAGACTT TGTGTCTACT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        65         75         85         95        105        115
EMCR5'UTR           CCTCTCAACT AAACGAAATT TTT-CTAGTG CTGTCATTTG TTATG--GCA GTCCTAGTGT
229E5'UTR           TTTCTCAACT AAACGAAATT TTTGCTATGG CCGGCATCTT TGATGCTGGA GTCGTAGTGT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       125        135        145        155        165        175
EMCR5'UTR           AATTGAAATT TCGTCAAGTT TGTAA-ACTG GTTAGGCAAG TGTTGTATTT TCTGTGTTTA
229E5'UTR           AATTGAAATT TCATTTGGGT TGCAACAGTT TGGAAGCAAG TGCTGTGTGT CCTA-GTCTA

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       185        195        205        215        225        235
EMCR5'UTR           AGCACTGGTG GTTCTGTC-C ACTAGTGCAC AC-ATTGATA CTTAAGT-GG TGTTCTGTCA
229E5'UTR           AGGGTTTCGT GTTCCGTCAC GAGATTCCAT TCTACAAACG CCTTACTCGA GGTTCCGTCT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....
                       245        255        265        275        285
EMCR5'UTR           CTGCTTATTG TGGAAGCAAC GTTCTGTCGT TGTGGAAACC AATAACTGCT AACC
229E5'UTR           CGTGTTTGTG TGGAAGCAAA GTTCTGTCTT TGTGGAAACC AGTAACTGTT CCTA
```

b. Putative Orf 1a

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         5         15         25         35         45         55
EMCR                ---------- ---MFYNQVT LAVASDSEIS GFGFAIPSVA VRAYSEAAAQ GFQACRFVAF
229E                ---------- ---MACNRVT LAVASDSEIS ANGCSTIAQA VRRYSEAASN GFRACRFVSL
PEDV                ---------- ---MASNHVT LAFANDAEIS AFGFCTASEA VSYYSEAAAS GFMQCRFVSL
TGEV                ---------- ---MSSKQFK ILVNEDYQVN VPSLPIR-DV LQEIKYCYRN GFEGYVFVPE
OC43                MSKINKYGLE LHWAPEFPWM FEDAEEKLDN PSSSEVDMIC STTAQKLETD GICPENHVMV
BoCoV               MSKINKYGLE LHWAPEFPWM FEDAEEKLDN PSSSEVDIVC STTAQKLETG GICPENHVMV
MHV                 MAKMGKYGLG FKWAPEFPWM LPNASEKLGS PERSEEDGFC PSAAQEPKTK GKTLINHVRV
AIPV                ---------- ---MASSLKQ GVSPKPRDVI LVSKDIPEQL CDALFFYTSH NPKDYADAFA
SARS CoV            ---------- ---MESLVLG VNEKTHVQLS LPVLQVRDVL VRGFGDSVEE ALSEAREHLK

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        65         75         85         95        105        115
EMCR                GLQDCVTGIN DDD-YVIALT GTNQLCAKIL LFSDRPLNLR GWLIFSNSNY VLQDFDVVFG
229E                DLQDCIVGIA DDT-YVMGLH GNQTLFCNIM KFSDRPFMLH GWLVFSNSNY LLEEFDVVFG
PEDV                DLADTVEGLL PED-YVMVVI GTTKLSAYVD TFGSRPRNIC GWLLFSNCNY FLEELELTFG
TGEV                YCRDLVDCDR KDH-YVIGVL GNGVSDLKPV LLTEPSVMLQ GFIVRANCNG VLEDFDLKIA
OC43                DCRRLLKQEC CVQSSLIREI VMNASPYDLE VLLQDALQSR EAVLVTTPLG MSLEACYVRG
BoCoV               DCRRLLKQEC CVQSSLIREI VMNTRPYDLE VLLQDALQSR EAVLVTPPLG MSLEACYVRG
MHV                 DCSRLPALEC CVQSAIIRDI FVDEDPLNVE ASTMMALQFG SAVLVKPSKR LSIQAWAKLG
AIPV                VRQKFDRSLQ TGKQFKFETV CGLFLLKGVD KITPG----- VPAKVLKATS KLADLEDIFG
SARS CoV            NGTCGLVELE KGVLPQLEQP YVFIKRSDAL STNHGHKVVE LVAEMDGIQY GRSGITLGVL

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       125        135        145        155        165        175
EMCR                --HGAGSVVF VDKYMCGFDG KPVLPKNMWE FRDYFNDNTD S-IVIGGVTY QLAWDVIRKD
229E                K-RGGGNVTY TDQYLCGADG KPVMSEDLWQ FVDHFGENEE --IIINGHTY VCAWLTKRKP
PEDV                --RRGGNIVP VDQYMCGADG KPVLQESEWE YTDFFADSED GQLNIAGITY VKAWIVERSD
TGEV                --RTGRGAIY VDQYMCGADG KPVIEG---D FKDYFGDED- -IIEFEGEEY HCAWTTVRDE
OC43                C-NPKGWTMG LFRRRSVCNT GRCTVNKHVA YQLYMIDPAG ---VCLGAGQ FVGWVIPLAF
BoCoV               C-NPNGWTMG LFRRRSVCNT GRCAVNKHVA YQLYMIDPAG ---VCFGAGQ FVGWVIPLAF
MHV                 V-LPKTPAMG LFKRFCLCNT RECVCDAHVA FQLFTVQPDG ---VCLGNGR FIGWFVPVTA
AIPV                VSPLARKYRE LLKTACQWSL TVEALDVRAQ TLDEIFDPT- ---------- EILWLQVAAK
SARS CoV            VPHVGETPIA YRNVLLRKNG NKGAGGHSYG IDLKSYDLGD ELGTDPIEDY EQNWNTKHGS

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       185        195        205        215        225        235
EMCR                LSYEQQNVLA IESIHYLG-T TGHTLKSGCK LINAKPPKY- --SSKVVLSG EWNAVYKAFG
229E                LDYKRQNNLA IEEIEYVHGD ALHTLRNGSV LEMAKEVKT- --SSKVVLSD ALDKLYKVFG
PEDV                VSYASQNLTS IKSITYCS-T YEHTFLDGTA MKVARTPKI- --KKNVVLSE PLATIYREIG
TGEV                KPLNQQTLFT IQEIQYNL-D IPHKLPNCAT RHVAPPVKK- --NSKIVLSE DYKKLYDIFG
OC43                MPVQSRKFIV PWVMYLRKRG EKGAYNKDHG RGGFGH----- --VYDFKVED AYDQVHDEPK
```

```
BoCoV       MPVQSRKFIA PWVMYLRKCG EKGAYIKDYK RGGFEH---- --VYNFKVED AYDLVHDEPK
MHV         IPAYAKQWLQ PWSILLRKGG NKGSVTSGHF RRAVTMP--- --VYDFNVED ACEEVHLNPK
AIPV        IHVSSMAMRR LVGEVTAKVM DALGSNLSAL FQIVKQ---- --QIARIFQK ALAIFENVNE
SARS CoV    GALRELTREL NGGAVTRYVD NNFCGPDGYP LDCIKDFLAR AGKSMCTLSE QLDYIESKRG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  245        255        265        275        285        295
EMCR        SPFITNGISL LDIIVKPVFF NAFVKCNCGS ENWSVGAWDG YLSSCCGTPA KKLCVVPGNV
229E        SPVMTNGSNI LEAFTKPVFI SALVQCTCGT KSWSVGDWTG FKSSCCNVIS NKLCVVPGNV
PEDV        SPFVDNGSDA RSIIRRPVFL HAFVKCKCGS YHWTVGDWTS YVSTCCGFKC KPVLVASCSA
TGEV        SPFMGNGDCL SKCFDTLHFI AATLRCPCGS ESSGVGDWTG FKTACCGLSG KVKGVTLGDI
OC43        GKFSKKAYAL IRGYRGVKPL LYVDQYGCDY TGSLADGLEA YADKTLQEMK ALFPTWSQEL
BoCoV       GKFSKKAYAL IRGYRGVKPL LYVDQYGCDY TGGLADGLEA YADKTLQEMK ALFPIWSQEL
MHV         GKYSRKAYAL LKGYRGVKSI LFLDQYGCDY TGRLAKGLED YGDCTLEEMK ELFPVWCDSL
AIPV        LPQRIAALKM AFAKCARSIT VVVVERTLVV KEFAGTCLAS INGAVAKFFE ELPNGFMGSK
SARS CoV    VYCCRDHEHE IAWFTERSDK SYEHQTPFEI KS--AKKFDT FKGECPKFVF PLNSKVKVIQ

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  305        315        325        335        345        355
EMCR        VPGDVIITST DAGCGVKYYA GLVVKHITNI TGVSLWRVTA VHSDGMFVAT SSYDALLHRN
229E        KPGDAVITTQ QAGAGIKYFC GMTLKFVANI EGVSVWRVIA LQSVDCFVAS STFVEEEHVN
PEDV        MPGSVVVTRA GAGTGVKYYN NMFLRHVADI DGLAFWRILK VQSKDDLACS GKFLEHHEEG
TGEV        KPGDAVVTSM SAGKGVKFFA NCVLQYAGDV EGVSIWKVIK TFTVDETVCT PGFEGELN--
OC43        LFDVIVAWHV VRDP----RY VMRLQSAATI R---SVAYVA NPTEDLCDGS VVIKEPVHVY
BoCoV       PFDVTVAWHV VRDP----RY VMRLQSASTI R---SVAYVA NPTEDLCDGS VVIKEPVHVY
MHV         DNEVVVAWHV DRDP----RA VMRLQTLATI R---SIGYVG QPTEDLVDGD VVVREPAHLL
AIPV        IFTTLAFFKE AAVR------ -VVENIPNAP RGTKGFEVVG NAKGTQVVVR GMRNDLTLLD
SARS CoV    PRVEKKKTEG FMGRIRSVYP VASPQECNNM HLSTLMKCNH CDEVSWQTCD FLKATCEHCG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  365        375        385        395        405        415
EMCR        SLDPFCFDVN TLLSNQLRLA FLGASVTEDV KFAASTGVID ISAGMFGLYD DILTNNKPWF
229E        RMDTFCFNVR NSVTDECRLA MLGAEMTSNV RRQVASGVID ISTGWFDVYD DIFAESKPWF
PEDV        FTDPCYFLND SSLATKLKFD ILSGKFSDEV KQAIIAGHVV VGSALVDIVD DALG--QPWF
TGEV        --DFIKPESK SLVACSVKRA FITGDIDDAV HDCIITGKLD LSTNLFGNVG LLFKK-TPWF
OC43        ADDSIILRQY NLVDIMSHFY MEADTVVNAF YGVALKDCGF VMQFGYIDCE QDSCDFKGWI
BoCoV       ADDSIILRQH NLVDIMSCFY MEADAVVNAF YGVDLKDCGF VMQFGYIDCE QDLCDFKGWV
MHV         AANAIVKRLP RLVETMLYT- --DSSVTEFC YKTKLCDCGF ITQFGYVDCC GDACDFRGWV
AIPV        QKADIPVEPE GWSAILDGHL CYVFRSGDRF YAAPLSGNFA LSDVHCCERV VCLSDGVTPE
SARS CoV    -TENLVIEGP TTCGYLPTNA VVKMPCPACQ DPEIGPEHSV ADYHNHSNIE TRLR--KGGR

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  425        435        445        455        465        475
EMCR        VRKASGLFDA IWDAFVAAIK LVPTTTGGLV RFVKSIASTV LTVSNGVIIM CADVPDAFQP
229E        VRKAEDIFGP CWSALASALK QLKVTTGELV RFVKSICNSA VAVVGGTIQI LASVPEKFLN
PEDV        IRKLGDLASA PWEQLKAVVR GLGLLSDEVV LFGKRLSCAT LSIVNGVFEF LADVPEKLAA
TGEV        VQKCGALFVD AWKVVEELCG SLTLTYKQIY EVVASLCTSA FTIVNYKPTF VVPD-NRVKD
OC43        PGNMIDGFAC TTCGHVYEVG DLIAQSSGVL PVNPVLHTKS AAGYGG---- -FGCKDSFTL
BoCoV       PGNMIDGFAC TTCGHVYETG DLLAQSSGVL PVNPVLHTKS AAGYGG---- -FGCKDSFTL
MHV         PGNMMDGFLC PGCSKSYMPW ELEAQSSGVI PKGGVLFTQS TDTVN----- ----RESFKL
AIPV        INDGLILAAI YSSFSVSELV TALKKGEPFK FLGHKFVYAK DAAVS----- -------FTL
SARS CoV    TRCFGGCVFA YVGCYNKRAY WVPRASADIG SGHTGITGDN VETLN----- --EDLLEILS

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  485        495        505        515        525        535
EMCR        VYRTFTQAIC AAFDFSLDVF KIG---DVKF KRLGDYVLTE NALVRLTTEV VRGVRDARIK
229E        AFDVFVTAIQ TVFDCAVETC TIA---GKAF DKVFDYVLLD NALVKLVTTK LKGVRERGLN
PEDV        AVTVFVNFLN EFFESACDCL KVG---GKTF NKVGSYVLFD NALVKLVKAK ARGPRQAGIC
TGEV        LVDKCVKVLV KAFDVFTQII TIAGIEAKCF VLGAKYLLFN NALVKLVSVK ILGKKQKGLE
OC43        YGQTVVYFGG CVYWSPARNI WIP--ILKSS VKSYDSLVYT GVLGCKAIVK ETNLICKALY
BoCoV       YGQTVVYFGG CVYWSPARNI WIP--ILKSS VKSYDGLVYT GVVGCKAIVK ETNLICKALY
MHV         YGHAVVPFGS AVYWSPYPGM WLP--VIWSS VKSYADLTYT GVVGCKAIVQ ETDAICRSLY
AIPV        AKAATIADVL RLFQSARVIA EDVWS-SFTE KSFEFWKLAY GKVRNLEEFV KTYVCKAQMS
SARS CoV    RERVNININVG DFHLNEEVAI ILAS-FSAST SAFIDTIKSL DYKSFKTIVE SCGNYKVTKG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  545        555        565        575        585        595
EMCR        KAMFTKVVVG PTTEVKFSVI ELATVNLRLV DCAPVVCPKG KIVVIAGQAF FYSGGFYRFM
229E        KVKYATVVVG STEEVKSSRV ERSTAVLTIA NNYSKLFDEG YTVVIGDVAY FYSDGYFRLM
PEDV        EVRYTSLVVG STTKVVSKRV ENANVNLVVV DEDVTLNTTG RTVVVDGLAF FESDGFYRHL
TGEV        CAFFATSLVG ATVNVTPKRT ETATISLNKV DDVVAPG-EG YIVIVGDMAF YKSGEYYFMM
OC43        LDYVQHKCGN LHQRELLGVS DVWHKQLLLN RGVYKPLLEN IDYFNMRRAK FSLETFTVCA
BoCoV       LDYVQHKCGN LHQRELLGVS DVWHKQLLLN RGVYKPLLEN IDYFNMRRAK FSLETFTVCA
MHV         MDYVQHKCGN LEQRAILGLD DVYHRQLLVN RGDYSLLLEN VDLFVKRRAE FACK-FATCG
AIPV        IVILAAVLGE DIWHLVSQVI YKLGVLFTKV VDFCDKHWKG FCVQLKRAKL IVTETFCVLK
SARS CoV    KPVKGAWNIG QQRSVLTPLC GFPSQAAGVI RSIFARTLDA ANHSIPDLQR AAVTILDGIS

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  605        615        625        635        645        655
EMCR        VDSTTVLNDP VFTGELFYTI KFSGFKLDGF N--------H QFVNASSATD AIIAVELLLS
```

```
                 229E        ASPNSVLTTA VYKPLFAFNV NVMGTRPE-- ---------- KFPTTVTCEN LESAVLFVND
                 PEDV        ADADVVIEHP VYKSACELKP VFECDPIP-- D--------F PLPVAASVAE LCVQTDLLLK
                 TGEV        SSPNFVLTNN VFKAVKVPSY DIVYDVDNDT KSKMIAKLGS SFEYDGDIDA AIVKVNELLI
                 OC43        DGFMPFLLDD LVPRAYYLAV SGQAFCDY-- ---------- ADKLCHAVVS KSKELLDVSL
                 BoCoV       DGFMPFLLDD LVPRAYYLAV SGQAFCDY-- ---------- AGKICHAVVS KSKELLDVSV
                 MHV         DGLVPLLLDG LVPRSYYLIK SGQAFTSM-- ---------- MVNFSHEVTD MCMDMALLFM
                 AIPV        GVAQHCFQLL LDAIHSLYKS FKKCALGR-- ---------- ---IHGDLLF WKGGVHKIVQ
                 SARS CoV    EQSLRLVDAM VYTSDLLTNS VIIMAYVTG- ---------- --GLVQQTSQ WLSNLLGTTV

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  665        675        685        695        705        715
                 EMCR        DFKTAVFVYT CVVDGCSVIV RRDAT-FATH VCFKDCYSIW EQFCIDNCGE PWFLTDYNAI
                 229E        KITEFQLDYS IDVIDNEIIV KPNIS-LCVP LYVRDYVDKW DDFCRQYSNE SWFEDDYRAF
                 PEDV        NYNTPYKTYS CVVRGDKCCI TCTLQ-FKAP SYVEDAVN-F VDLCTKNIGT AGFHEFYITA
                 TGEV        EFRQQSLCFR AFKDDKSIFV EAYFKKYKMP ACLAKHIG-L WNIIKKDSCK RGFLNLFNHL
                 OC43        DSLGAAIHYL NSKIVDLAQH FSDFG----- TSFVSKIVHF FKTFTTSTAL AFAWVLFHVL
                 BoCoV       DSLGAAIHYL NSKIVDLAQH FSDFG----- TSFVSKIVHF FKTFTTSTAL AFAWVLFHVL
                 MHV         HDVKVATKYV KKVTGKLAVR FKALG----- VAVVRKITEW FDLAVDTAAS AAGWLCYQLV
                 AIPV        DGDEIWFDAI DSVDVEDLGV VQEKS---ID FEVCDDVTLP ENQPGHMVQI EDDGKNYMFF
                 SARS CoV    EKLRPIFEWI EAKLSAGVEF LKDAW----- EILKFLITGV FDIVKGQIQV ASDNIKDCVK

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  725        735        745        755        765        775
                 EMCR        LQSNNPQCAI VQASESK--V LLERFLPKCP EILLSIDDGH LWNLFVEKFN FVTDWLKTLK
                 229E        ISVLDITDAA VKAAESK--A FVDTIVPPCP SILKVIDGGK IWNGVIKNVN SVRDWLKSLK
                 PEDV        HEQQDLQGFL TTCCTMSGFE CFMPTIPQCP AVLEEIDGGS IWRSFITGLN TMWDFCKRLK
                 TGEV        NELEDIKETN IQAIKN---- ------ILCP DPLLDLDYGA IWYNCMPGCS DP-SVLGSVQ
                 OC43        HGAYIVVESD IYFVKN---- -IPRYASAVA QAFQSVAKVV LDSLRVTFID GLSCFKIGRR
                 BoCoV       HGAYIVVESD IYFGKN---- -IPRYASAVA QAFRSGAKVG LDSLRVTFID GLSCFKIGRR
                 MHV         NGLFAVANGG ITFLSD---- -VPELVKNFV DKFKVFFKVL IDSMSVSVLS GLTVVKTASN
                 AIPV        RFKKDENIYY TPMSQLG--- ---AINVVCK AGGKTVTFG- --ETTVQEIP PPDVVPIKVS
                 SARS CoV    CFIDVVNKAL EMCIDQ---- -----VTIAG AKLRSLNLGE VFIAQSKGLY RQCIRGKEQL

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  785        795        805        815        825        835
                 EMCR        LTLTSNGLLG NCAKRFRRVL VKLLDVYNGF LETVCSVVHT AGVCIKYYAV NVP-YVVISG
                 229E        LNLTQQGLLG TCAKRFKRWL GILLEAYNAF LDTVVSTVKI GGLTFKTYAF DKP-YIVIRD
                 PEDV        VSFGLDGIVV TVARKFKRLG ALLAEMYNTY LSTVVENLVL AGVSFKYYAT SVP-KIVLGG
                 TGEV        LLIGNG--VK VVCDGCKGFA NQLSKGYNKL CNAARNDIEI GGIPFSTFKT PTNTFIEMTD
                 OC43        RICLSGRKIY EVERGLLHSS QLPLDVYDLT MPSQVQKAKQ KPIYLKGSGS DFSLADSVVE
                 BoCoV       RICLSGSKIY EVERGLLHSS QLPLDVYDLT MPSQVQKTKQ KGIYLKGSGS DFSLADSVVE
                 MHV         RVCLAGCKVY EVVQKRLSAY VMPVGCNEAT C---------- ----LVGEIE PAVVEDDVVD
                 AIPV        IECCGEPWNT IFKKAYKEPI EVDTDLTVEQ LLSVIYEKMC DDLKLFPEAP EPPPFENVAL
                 SARS CoV    QLLMPLKAPK EVTFLEGDSH DTVLTSEEVV LKNGELEALE TPVDSFTNGA IVGTPVCVNG

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  845        855        865        875        885        895
                 EMCR        FVSRVIRRER CD--VTFPCV SCVTFFYEFL DTCFGVSK-- --PNAIDVEH LELKETVFVE
                 229E        IVCKVENKTE AEWIELFPHN DRIKSFSTFE SAYMPIAD-- --PTHFDIEE VELLDAEFVE
                 PEDV        CFHSVKSVFA SV--FQIPVQ AGIEKFKVFL NCVHPVV--- --PRVIETSF VELEETTFKP
                 TGEV        AIYSVIEQGK AL-------- S--------FR DADVPVVDNG TISTADWSEP ILLEPAEYVK
                 OC43        VVTTSLTPCG YS------EPP KVADKICIVD NVYMAKAGDK YYPVVVD-DH VGLLDQAWRV
                 BoCoV       VVTTSLTPCG YS------EPP KVADKICIVD NVYMAKAGDK YYPVVVD-GH VGLLDQAWRV
                 MHV         VVKAPLTYQG CC------KPP TSFEKICVVD KLYMAKCGDQ FYPVVVDNDT IGVLDQCWRF
                 AIPV        VDKNGKDLDC IKS-------- ------CHLI YRDYESD--- --------DD IEEEDAEECD
                 SARS CoV    LMLLEIKDKE QY-------- ------CALS PGLLATN--- --------NV FRLKGGAPIK

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  905        915        925        935        945        955
                 EMCR        PKDGGQFFVS DDYLWYVV-D D-------IY YPASCNGVLP VAFTKLAGGK ---ISFSDDV
                 229E        PGCGGILAVI DEHVFYKK-D G--------VY YPSNGTNILP VAFTKAAGGK ---VSFSDDV
                 PEDV        PALNGGIAIV DGFAFYYD-G T--------LY YPTDGNSVVP ICFKKKGGGD ---VKFSDEV
                 TGEV        PKNNGNVIVI AGYTFYKDED E--------HF YPYGFGKIVQ RMYNKMGGGD KT-VSFSEEV
                 OC43        PCAGRRVTFK EQPTVKEIIS MPKIIKVFYE LDNDFNTILN TACGVFEVDD TVDMEEFYAV
                 BoCoV       PCAGRCVTFK EQPTVNEIAS TPKTIKVFYE LDKDFNTILN TACGEFEVDD TVDMEEFYAV
                 MHV         PCAGKKVEFN DKPKVKEIPS T-RKIKINFA LDATFDSVLS KACSEFEVDK DVTLDELLDV
                 AIPV        TDSGEAEECD TNSECEEEDE D--------TK VLALIQDPAS IKYPLPLDED YS-VYNGCIV
                 SARS CoV    GVTFGEDTVW EVQGYKNVRI T--------FE LDERVDKVLN EKCSVYTVES GTEVTEFACV

                             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  965        975        985        995       1005       1015
                 EMCR        IVHDVEPTHK VKLIFEFEDD -VVTSLCKKS FGKSIIYTG- DWEGLHEVLT SAMNVIG---
                 229E        EVKDIEPVYR VKLCFEFEDE -KLVDVCEKA IGKKIKHEG- DWDSFCKTIQ SALSVVS---
                 PEDV        SVKTIDPVYK VSLEFEFESE -TIMAVLNKA VGNRIKVTG- GWDDVVEYIN VAIEVLK---
                 TGEV        DVQEIAPVTR VKLEFEFDNE -IVTGVLERA IGTRYKFTGT TWEEFEESIS EELDAIFDTL
                 OC43        VIDAIEEKLS PCKELEGVGA -KVSAFLQKL EDNPLFLFD- --EAGEEVLA PKLYCAFTAP
                 BoCoV       VIDAIEEKLS PCKELEGVGA -KVSAFLQKL EDNSLFLFD- --EAGEEVLA PKLYCAFTAP
                 MHV         VLDAVESTLS PCKEHDVIGT -KVCALLNRL AEDYVYLFD- --EGGEEVIA PKMYCSFSAP
                 AIPV        HKDALDVVNL PSGEETFVVN NCFEGAVKPL PQKVVDVLG- --DWGEAVDA QEQLCQQ---
                 SARS CoV    VAEAVVKTLQ PVSDLLTN-- --MGIDLDEW SVATFYLFD- --DAGEENFS SRMYCSFYPP
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1025       1035       1045       1055       1065       1075
EMCR                --QHIKLPQF YIYDEEGGYD VSKP--VMIS QWPISDDSDG CVVEASTDFH Q--LESVREE
229E                --CYVNLPTY YIYDEEGGND LSLP--VMIS EWPLSVQQAQ QEATLPDIAE D--VVDQVEE
PEDV                --DHVEVPKY YIYDEEGGTD PNLP--VMVS QWPLNDDTIS QDLLDVEVVT DAPIDSEGDE
TGEV                ANQGVELEGY FIYDTCGGFD IKNPDGIMIS QYDINITADE KSEVSASSEE EE-VESVEED
OC43                EDDDFLEESD VEEDDVEGEE TDLTVTSAGQ PCVASEQEES SEVLEDTLDD GPSVETSDSQ
BoCoV               EDDDFLEESG VEEDDVEGEE TDLTVTSAGE PCVASEQEES SEILEDTLDD GPCVETSDSQ
MHV                 DDEDCVAADV VDADENQGDD ADDSAALVTD TQEEDGVAKG QVGVAESDAR LDQVEAFDIE
AIPV                --EPLQHTFE EPVENSTGSS KTMTEQVVVE DQELPVVEQD QDVVVYTPTD LEVAKETAEE
SARS CoV            DEEEEDDAEC EEEEIDETCE HEYGTEDDYQ GLPLEFGASA ETVRVEEEEE EDWLDDTTEQ

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1085       1095       1105       1115       1125       1135
EMCR                VD--------  ----------  ----------  -------IIE QPFGEVEHAL SIRQ------
229E                VNS-------  ----------  ----------  -------IFD IETVDVKHDV S---------
PEDV                VDSSAPEKVA  ----------D VANSEPGDDG LPVAPETNVE SEVEEVAATL SFIKDTPSTV
TGEV                PENEIVEASE GAEGTSSQEE VETVEVADIT STEEDVDIVE VSAKDDPWAA AVDVQEAEQF
OC43                VEEDVEMS-- -----DFVDL ESVIQD---- ---------- --YENVCFEF YTT-------
BoCoV               VEEDVQMS-- -----DFGDL ESVIQD---- ---------- --YENVCFEF YTT-------
MHV                 KVEDPILN-- -----ELSAE LNAPADK--- -TYEDVLAFD AIYSEALSAF YAVP------
AIPV                VD-------- ---------- ---------- ---------- ---------- ----------
SARS CoV            SEIEPEP--- ---------- ---------- ----------E PTPEEPVNQF TG--------

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1145       1155       1165       1175       1185       1195
EMCR                ---PFSFSFR DELGVRVLDQ SDNNCWISTT LIQLQLTKLL DDSIEMQLFK VGKVDSIVQK
229E                ---PFEMPFE ELNGLKILKQ LDNNCWVNSV MLQIQLTGIL DGDYAMQFFK MGRVAKMIER
PEDV                TKDPFAFDFV SYGGLKVLRQ SHNNCWVTST LVQLQLLGIV DDP-AMELFS AGRVGPMVRK
TGEV                NPSLPPFKTT NLNGKIILKQ GDNNCWINAC CYQLQAFDFF NNE-AWEKFK KGDVMDFVNL
OC43                -----EPEFV KVLGLYVPKA TRNNCWLRSV LAVMQKLPCQ FKD--KNLQD LWVLYKQQYS
BoCoV               -----EPEFV KVLDLYVPKA TRNNCWLRSV LAVMQKLPCQ FKD--KNLQD LWVLYKQQYS
MHV                 ----GDETHF KVCGFYSPAI ERTNCWLRST LIVMQSLPLE FKD--LEMQK LWLSYKSSYN
AIPV                ---------- EFILIFAVPK EEVVSQKDGA QIKQEPIQVV KPQ--REKKA KKFKVKPATC
SARS CoV            ----YLKLTD NVAIKCVDIV KEAQSANPMV IVNAANIHLK HGGGVAGALN KATNGAMQKE

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1205       1215       1225       1235       1245       1255
EMCR                CYELSHLISG SLGDSGKLLS ELLKDKYTCS ITTEMSCDCG KKFDEQVGCL FWIMPYTKLF
229E                CYTAEQCIRG AMGDVGLCMY RLLKDLHTGF MVMDYKCSCT SGRLEESGAV LFCTPTKKAF
PEDV                CYESQKAILG SLGDVSACLE SLTKDLHTLK ITCSVVCGCG TGERIYEGCA FRMTPTLEPF
TGEV                CYAATTLARG HSGDAEYLLE LMLNDYSTAK IVLAAKCGCG EKEIVLERAV FKLTPLKESF
OC43                QLFVDTLVNK IPANIVLPQG GYVADFAYWF LTLCDWQCVA YWKCIKCDLA LKLKGLDAMF
BoCoV               QLFVDTLVNK IPANIVVPQG GYVADFAYWF LTLCDWQCVA YWKCIKCDLA LKLKGLDAMF
MHV                 KEFVDKLVKS VPKSIILPQG GYVADFAYFF LSQCSFKAYA NWRCLKCDMD LKLQGLDAMF
AIPV                EKPKFLEYKT CVGDLTVVIA KALDEFKEFC IVNAANEHMT HGSGVAKAIA DFCGLDFVEY
SARS CoV            SDDYIKLNGP LTVGGSCLLS GHNLAKKCLH VVGPNLNAGE DIQLLKAAYE NFNSQDILLA

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1265       1275       1285       1295       1305       1315
EMCR                QKGECCICHK MQTYKLVSMK GTGVFVQD-- PAPIDIDAFP VRPICSSVYL GVKGSGHYQT
229E                PYGTCLNCNA PRMCTIRQLQ GTIIFVQQK- PEPVNPVSFV VKPVCSSIFR GAVSCGHYQT
PEDV                PYGACAQCAQ VLMHTFKSIV GTGIFCRD-- TTALSLDSLV VKPLCAAAFI GK-DSGHYVT
TGEV                NYGVCGDCMQ VNTCRFLSVE GSGVFVHDIL SKQTPEAMFV VKPVMHAVYT GTTQNGHYMV
OC43                FYGDVVSHIC KCGESMVLID VDVPFTAHFA LKDKLFCAFI TKRIVYKAAC VVDVNDSHSM
BoCoV               FYGDVVSHVC KCGESMVLID VDVPFTAHFA LKDKLFCAFI TKRSVYKAAC VVDVNDSHSM
MHV                 FYGDVVSHVC KCGTGMTLLS ADIPYTLHFG LRDDKFCAFY TPRKVFRAAC VVDVNDCHSM
AIPV                CEDYVKKHGP QQRLVTPSFV KGIQCVNNVV GPRHGDNNLH EKLVAAYKNV LVDGVVNYVV
SARS CoV            PLLSAGIFGA KPLQSLQVCV QTVRTQVYIA VNDKALYEQV VMDYLDNLKP RVEAPKQEEP

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1325       1335       1345       1355       1365       1375
EMCR                NLYSFDKAID GFGVFDIK-- ---------- ------NSSV NTVCFVDVDF HS-VEIEAGE
229E                NIYSQNLCVD GFGVNKIQP- ---------- ----WTNDAL NTICIKDADY NAKVEISVTP
PEDV                NFYDAAMAID GYGRHQIK-- ---------- ------YDTL NTICVKDVNW TAPLVPAVDS
TGEV                DDIEHGYCVD GMGIKPLKKR CYTSTLFINA NVMTRAEKPK QEFKVEKVEQ QPIVEENKSS
OC43                AVVDG-KQID DHRITSIT-- ---------- -------SDK FDFIIGHGMS FSMTTFEIAQ
BoCoV               AVVDG-KQID DHRITSIT-- ---------- -------SDK FDFIIGHGTS FSMTTFEIAQ
MHV                 AVVDG-KQID GKVVTKFN-- ---------- -------GDK YDFMVGHGMA FSMSAFEIAQ
AIPV                PVLSLGIFGV DFKMSIDAMR ---------- --------EA FEGCTIRVLL FSLSQEHIDY
SARS CoV            PNTEDSKTEE KSVVQKPVDV KP-------- --KIKACIDE VTTTLEETKF LTNKLLLFAD

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      1385       1395       1405       1415       1425       1435
EMCR                VK-------- ---------- PFAVYKNVKF YLGDISHLVN CVSFDFVVNA ANENLMHGGG
229E                IKNTVDTTPK EEFVVKEKLN AFLVHDNVAF YQGDVDTVVN GVDFDFIVNA ANENLAHGGG
PEDV                VVEP------ -------VVK PFYSYKNVDF YQGDFSDLVK -LPCDFVVNA ANEKLSHGGG
TGEV                IEKEEIQSPK ND----DLIL PFYKAGKLSF YQGALDVLIN FLEPDVIVNA ANGDLKHMGG
OC43                LYG------- ---------- -SCITPNVCF VKGDIIKVSK LVKAEVVVNP ANGHMAHGGG
```

```
                BoCoV       LYG------- ---------- -SCITPNVCF VKGDIIKVSK RVKAEVVVNP ANGHMAHGGG
                MHV         LYG------- ---------- -SCITPNVCF VKGDVIKVLR RVGAEVIVNP ANGRMAHGAG
                AIPV        FD-------- ---------- -------VTC KQKTIYLTED GVKYRSIVLK PGDSLGQFGQ
                SARS CoV    INGKLYHD-- --------SQ NMLRGEDMSF LEKDAPYMVG DVITSGDITC VVIPSKKAGG
5
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1445       1455       1465       1475       1485       1495
                EMCR        VARAIDILTE GQLQSLSKDY ISSNGPLKVG AGVMLE--CE KFNVFNVVGP RTG---KHEH
                229E        LAKALDVYTK GKLQRLSKEH IGLAGKVKVG TGVMVE--CD SLRIFNVVGP RKG---KHER
10              PEDV        IAKAIDVYTK GMLQKCSNDY IKAHGPIKVG RGVMLE--AL GLKVFNVVGP RKG---KHAP
                TGEV        VARAIDVFTG GKLTERSKDY LKKNKSIAPG NAVFFENVIE HLSVLNAVGP RNGD--SRVE
                OC43        VAKAIAVAAG QQFVKETTDM VKSKGVCATG DCYVSTGGKL CKTVLNVVGP DARTQGKQSY
                BoCoV       VAKAIAVAAG QQFVKETTDM VKSKGVCATG DCYVSTGGKL CKTVLNVVGP DARTQGKQSY
                MHV         VAGAIAKAAG KSFIKETADM VKNQGVCQVG ECYESTGGNL CKTVLNIVGP DARGHGKQCY
15              AIPV        VYAKNKIVFT ADDVEDKEIL YVPTTDKSIL EYYGLD---A QKYVIYLQTL AQKWNVQYRD
                SARS CoV    TTEMLSRALK KVPVDEYITT YPGQGCAGYT LEEAKTALKK CKSAFYVLPS EAPNAKEEIL

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1505       1515       1525       1535       1545       1555
20              EMCR        SLLVEAYNSI LF---ENGIP LMPLLSCGIF GVRIENSLKA LFSCDINKPL QVFVYSSNEE
                229E        DLLIKAYNTI NN---EQGTP LTPILSCGIF GIKLETSLEV LLDVCNTKEV KVFVYTDTEV
                PEDV        ELLVKAYKSV FA---NSGVA LTPLISVGIF SVPLEESLSA FLACVGDRHC KCFCYGDKER
                TGEV        AKLCNVYKAI AK---CEGKI LTPLISVGIF NVRLETSLQC LLKTVNDRGL NVFVYTDQER
                OC43        VLLERVYKHL N-----NYDCV VTTLISAGIF SVPSDVSLTY LLGTAKKQVV LVSNNQEDFD
25              BoCoV       ALLERVYKHL N-----KYDCV VTTLISAGIF SVPSDVSLTY LLGTAKKQVV LVSNNQEDFD
                MHV         SFLERAYQHI N-----KCDDV VTTLISAGIF SVPTDVSLTY LIGVVTKNVI LVSNNKDDFD
                AIPV        NFLILEWRDG N-----CWISS AIVLLQAAKI RFKGFLTEAW AKLLGGDPTD FVAWCYASCT
                SARS CoV    GTVSWNLREM LAHAEETRKL MPICMDVRAI MATIQRKYKG IKIQEGIVDY GVRFFFYTSK
30
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1565       1575       1585       1595       1605       1615
                EMCR        QAVLKFLDGL DLTPVID--- ------DVDVV ---------- -KPFRVEGNF SFFDCG---V
                229E        CKVKDFVSGL VNVQKVE--- ------QPKIE PKPVSVIKVA PKPYRVDGKF SYFTED---L
                PEDV        EAIIKYMDGL VDAIFKEALV DTTPVQEDVQ QVSQKPVLPN FEPFRIEGAH AFYECNPEGL
35              TGEV        QTIENFFS-- ---------- ---------- ---------- ---------- ----------
                OC43        LISKCQITAV EG-------- ---------- ---------- ---------- ----------T
                BoCoV       LISKCQITAV EG-------- ---------- ---------- ---------- ----------T
                MHV         VIEKCQVTSI AG-------- ---------- ---------- ---------- ----------T
                AIPV        AKVGDFSDAN ---------- ---------- ---------- ---------- ----------
40              SARS CoV    EPVASIITKL N--------- ---------- ---------- ---------- ----------S

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1625       1635       1645       1655       1665       1675
                EMCR        NALDGD-IYL LFTNSILMLD KQGQLLDTKL NGILQQAVLD YLATVKTVPA GNLVKLVVE-
45              229E        LCVADDKPIV LFTDSMLTLD DRGLALDNAL SGVLSAAIKD CVDINKAIPS GNLIKFDIG-
                PEDV        MSLGAD-KLV LFTNSNLDFC SVGKCLNDVT SGALLEAINV FKKSNKTVPA GNCVTLDCAN
                TGEV        ---------- ---------- ---------- ---------- ---------- ----------
                OC43        KKLAARLSFN VGRSIVYETD ANKLILIN-- -DVAFVSTFN VLQDVLSLRH DIALDDDART
                BoCoV       KKLAERLSFN VGRSIVYETD ANKLILSN-- -DVAFVSTFN VLQDVLSLRH DIALDDDART
50              MHV         KALSLQLAKN LCRDVKFETN ACDSLFS--- -DSCFVSSYD VLQEVELLRH DIQLDDDARV
                AIPV        ---------- ---------- ---------- ---------- ------WLLA NLAEHFDADY
                SARS CoV    LNEPLVTMPI GYVTHGFNLE EAARCMR--- -SLKAPAVVS VSSPDAVTTY NGYLTSSSKT

55                              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1685       1695       1705       1715       1725       1735
                EMCR        SCTIYMCVVP SI-NDLSFDK NLGRCVRKLN RLKTCVIANV PAIDVLKKLL SSLTLTVKFV
                229E        SVVVYMCVVP SE-KDKHLDN NVQRCTRKLN RLMCDIVCTI PADYILPLVL SSLTCNVSFV
                PEDV        MISITMVVLP FD-GDANYDK NYARAVVKVS KLKGKLVLAV DDATLYSKLS --HLSVLGFV
                TGEV        ------CSIP ---------- ---------- ---------- ---------- ----------
60              OC43        FVQSNVDVVP EG-WRVVNKF YQINGVRTVK YFECTGGIDI CSQDKVFGYV QQGIFNKATV
                BoCoV       FVQSNVDVVP EG-WRVVNKF YQINGVRPVK YFECPGGIDI CSQDKVFGYV QQGSFNKATV
                MHV         FVQAHMDNLP AD-WRLVNKF DSVDGVRTVK YFECPGEIFV SSQGKKFGYV QNGSFKVASV
                AIPV        TNAFLKKRVS CN-------- ---------- -----CG--- ---------- ----------
                SARS CoV    SEEHFVETVS LAGSYRDWSY SGQRTELGVE FLKRGDKIVY HTLESPVEFH LDG--EVLSL
65
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1745       1755       1765       1775       1785       1795
                EMCR        VESNVMDVND CFKNDNVVLK ITEDGINVKD VVVESSKSLG KQLG-VVSDG VDSFEGVLP-
                229E        GELKAAEA-- -----KVITIK VTEDGVNVHD VTVTTDKSFE QQVG-VIADK DKDLSGAVPS
70              PEDV        STPDDVER-- FYANKSVVIK VTEDTRSVKA VKVESTATYG QQIG-PCLVN DTVVTDNKP-
                TGEV        ---------- ---------VN VTEDNVNHER VSVSFDKTYG EQLKGTVVIK DKDVTNQLPS
                OC43        AQIKALFLD- -----KVDIL LTVDGVNFTN RFVPVGESFG KSLG-NVFCD GVNVTKHKCD
                BoCoV       AQIKALFLD- -----KVDIL LTVDGVNFTN RFVPVGESFG KSLG-NVFCD GVNVTKHKCD
                MHV         SQIRALLAN- -----KVDVL CTVDGVNFRS CCVAEGEVFG KTLG-SVFCD GINVTKVRCS
75              AIPV        ---------- -----IKSYE LRGLEACIQP VRATNLLHFK TQYSNCPTCG ANNTDEVIEA
                SARS CoV    DKLKSLLSLR --EVKTIKVF TTVDNTNLHT QLVDMSMTYG QQFG-PTYLD GADVTKIKPH

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    1805       1815       1825       1835       1845       1855
80              EMCR        -INTDTVLSV APEVDWVAFY GFEKAALFAS LDVKPYG--- -----YPNDF VGGFRVLGTT
```

```
229E        DLNTSELLTK AIDVDWVEFY GFKDAVTFAT VDHSAFA--- -----YESAV VNGIRVLKTS
PEDV        -VVADVVAKV VPNANWDSHY GFDKAGEFHM LDHTGFT--- -----FPSEV VNGRRVIKTT
TGEV        AFDVGQKVIK AIDIDWQAHY GFRDAAAFSA SSHDAYK--- -----FEVVT HSNFIVHKQT
OC43        INYKGKVFFQ FDNLSSEDLK AVRSSFNFDQ KELLAYYNML VNCFKWQVVV NGKYFTFKQA
BoCoV       INYKGKVFFQ FDNLSSEDLK AVRSSFNFDQ KELLAYYNML VNCSKWQVVF NGKYFTFKQA
MHV         AIHKGKVFFQ YSGLSAADLV AVTDAFGFDE PQLLKYYNML G-MCKWPVVV CGNYFAFKQS
AIPV        SLPYLLLFAT DGPATVDCDE DAVGTVVFVG STNSGHCY-- -------TQA AGQAFDNLAK
SARS CoV    VNHEGKTFFV LPSDDTLRSE AFEYYHTLDE SFLGRYMSAL NHTKKWKFPQ VGGLTSIKWA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1865       1875       1885       1895       1905       1915
EMCR        DNNCWVNATC IILQYLKPTF KSKGLNVLWN KFVTGDVGPF VSFIYFITMS SKGQKGDAEE
229E        DNNCWVNAVC IALQYSKPHF ISQGLDAAWN KFVLGDVEIF VAFVYYVARL MKGDKGDAED
PEDV        DNNCWVNVTC LQLQFARFRF KSAGLQAMWE SYCTGDVAMF VHWLYWLTGV DKGQPSDSEN
TGEV        DNNCWINAIC LALQRLKPQW KFPGVRGLWN EFLERKTQGF VHMLYHISGV KKGEPGDAEL
OC43        NNNCFVNVSC LMLQSLHLTF KIVQWQEAWL EFRSGRPARF VALVLAKGGF KFGDPADSRD
BoCoV       NNNCFVNVSC LMLQSLNLKF KIVQWQEAWL EFRSGRPARF VSLVLAKGGF KFGDPADSRD
MHV         NNNCYINVAC LMLQHLSLKF HKWQWQEAWN EFRSGKPLRF VSLVLAKGSF KFNEPSDSTD
AIPV        DRKFGKKSPY ITAMYTRFAF KN-ETSLPVA KQSKGKSKSV KEDVSNLATS SKASFDNLTD
SARS CoV    DNNCYLSSVL LALQQLEVKF NAPALQEAYY RARAGDAANF CALILAYSNK TVGELGDVRE

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1925       1935       1945       1955       1965       1975
EMCR        ALSKLSEYLI S---------- ---DSIVTLE QYSTCDIC-- ---------- ----------
229E        TLTKLSKYLA N---------- ---EAQVQLE HYSSCVECDA K---------- ----------
PEDV        ALNMLSKYIV P---------- ---AGSVTIE RVTHDGCC-- ---------- ----------
TGEV        MLHKLGDLMD N---------- ---DCEIIVT HTTACDKC-- ---------- ----------
OC43        FLRVVFSQVD LTGAICDF-E IACKCGVKQE QRTGLDAVMH FGTLSREDLE IGYTVDCSCG
BoCoV       FLRVVFSQVD LTGAICDF-E IACKCGVKQE QRTGVDAVMH FGTLSREDLE IGYTVDCSCG
MHV         FMRVVLREAD LSGATCDF-E FVCKCGVKQE QRKGVDAVMH FGTLDKGDLA KGYTIACTCG
AIPV        FEQWYDSNIY ES-------- -----LKVQE SPDNFDKY-- ---------- ----------
SARS CoV    TMTHLLQHAN LESAKRVLNV VCKHCGQKTT TLTGVEAVMY MGTLSYDNLK TGVSIPCVCG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1985       1995       2005       2015       2025       2035
EMCR        ---------- ---------- KSTVVEVKSA VVCASVLKDG ---------- CDVGFCPHRH
229E        ---------- ---------F KNSVASINSA IVCASVKRDG ---------- VQVGYCVHGI
PEDV        ---------- ---------- -CSKRVVTAP VVNASVLKLG ---------- VEDGLCPHGL
TGEV        ---------- ---------- -AKVEKFVGP VVAAPLAIHG ---------- TDE-TCVHGV
OC43        KKLIHCVRFD VP--FLICSN TPASVKLPKG VGSANIFIGD KVG-HYVHVK CEQSYQLYDA
BoCoV       KKLIHCVRFD VP--FLICSN TPASVKLPKG VGSANIFKGD KVG-HYVHVK CEQSYQLYDA
MHV         NKLVHCTQLN VP--FLICSN KPEGKKLPDD VVAANIFTGG SLG-HYTHVK CKPKYQLYDA
AIPV        ---------- ---------- -VSFTTKEDS KLPLTLKVRG ---------IK SVVDFRSKDG
SARS CoV    RDATQYLVQQ ESSFVMMSAP PAEYKLQQGT FLCANEYTGN YQCGHYTHIT AKETLYRIDG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  2045       2055       2065       2075       2085       2095
EMCR        KLRSRVKFVN G---------- ---------- ---------- ---------- -RVVITNVGE
229E        KYYSRVRSVR G---------- ---------- ---------- ---------- -RAIIVSVEQ
PEDV        NYIGKVVVVK G---------- ---------- ---------- ---------- -TTIVVNVGK
TGEV        SVNVKVTQIK G---------- ---------- ---------- ---------- -TVAITSLIG
OC43        SNVKKVTDVT GKLSDCLYLK NLKQTFKSVL TTYYLDDVKK IEYKPDLSQY YCDGGKYYTQ
BoCoV       SNVKKVTDVT GNLSDCLYLK NLKQTFKSVL TTYYLDDVKK IEYKPDLSQY YCDGGKYYTQ
MHV         CNVSKVSEAK GNFTDCLYLK NLKQTFSSKL TTFYLDDVKC VEYNPDLSQY YCESGKYYTK
AIPV        FIYKLTPDTD EN-------- ---------- ---------- ----S----- -KAPVYYPVL
SARS CoV    AHLTKMSEYK GPVTDVFYKE TSYTTTIKPV SYKLDGVTYT EIEPKLDGYY KKDNAYYTEQ

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  2105       2115       2125       2135       2145       2155
EMCR        PIISQPSKLL NGIA--YTTF S---------- ------GSFD NGHYVVYDAA NNAVYDGARL
229E        LEPCAQSRLL SGVA--YTAF S---------- ------GPVD KGHYTVYDTA KKSMYDGDRF
PEDV        PVVAPSHLFL KGVS--YTTF LDN------- ------GNGV VGHYTVFDHG TGMVHDGDAF
TGEV        PIIG---EVL EATG--YICY S---------- ------GSNR NGHYTYYDNR NGLVVDAEKA
OC43        RIIKAQFKTF EKVDGVYTNF KLIGHTVCDS LNA-KLGFDS SKEFVEYKIT EWPTATGDVV
BoCoV       RIIKAQFKTF EKVDGVYTNF KLIGHTVCDI LNA-KLGFDS SKEFVEYKVT EWPTATGDVV
MHV         PIIKAQFRTF EKVEGVYTNF KLVGHSIAEK FNA-KLGFDC NSPFTEYKIT EWPTATGDVV
AIPV        DAISLKAIWV EGNANFVVGH PN--------- -----YYSKS LHIPTFWENA ENFVKMGDKI
SARS CoV    PIDLVPTQPL PNAS--FDNF KLTCSNTKFA DDLNQMTGFT KPASRELSVT FFPDLNGDVV

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  2165       2175       2185       2195       2205       2215
EMCR        FASD------ ---------- ---------- ---------- --LSTLAVTA IVVVGGCVTS
229E        VKHD------ ---------- ---------- ---------- --LSLLSVTS VVMVGGYVA-
PEDV        VPGD------ ---------- ---------- ---------- --LNVSPVTN VVVSEQTAVV
TGEV        YHFN------ ---------- ---------- ---------- --RDLLQVTT AIASNFVVKK
OC43        LATDDLYVKR YERGCITFGK PVIWLS---- ----HEKASL NSLTYFNRPS LVDDNKFDVL
BoCoV       LATDDLYVKR YERGCITFGK PVIWLS---- ----HEQASL NSLTYFNRPL LVDENKFDVL
MHV         LASDDLYVSR YSGGCVTFGK PVIWLG---- ----HEEASL KSLTYFNRPS VVCENKFNVL
AIPV        GGVT------ ---------- ---------- ---------- --MGLWRAEH LNKPNLERIF
SARS CoV    AIDYRHYSAS FKKGAKLLHK PIVWHINQAT TKTTFKPNTW CLRCLWSTKP VDTSNSFEVL
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2225       2235       2245       2255       2265       2275
EMCR                ---------- ---------- ---------- ---------- ---------- ----------
229E                ---------- ---------- ---------- ---------- ---------- ----------
PEDV                ---------- ---------- ---------- ---------- ---------- ----------
TGEV                PQAEER---- ---------- ---------- ---------- ---------- --------P--
OC43                KVDDVD---- ---------- ---DGGDSSE SGAKE----- ---------- ---------T
BoCoV               KVDDVD---- ---------- ---DGGDISE SDAKE----- ---------- ---------P
MHV                 PVDVSEPTDK GPVPAAVLVT GALSGAATAP GTAKEQKVCA SDSVVDQVVS GFLSDLSGAT
AIPV                N--------- ---------- ---------- ---------- ---------- ----------
SARS CoV            AVEDTQG--- ---------- MDNLACESQQ PTSEEVVEN- ---------- ----------

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2285       2295       2305       2315       2325       2335
EMCR                ---------- ---------- ------NVPP IVSEKISVMD KLDTG----- ----------
229E                ---------- ---------- --------PV NTVKPKPVIN QLDEK----- ----------
PEDV                ---------- ---------- ------IKDP VKKAELDATK LLDTMNY--- ----------
TGEV                ---------- --------KN CAFNKVAASP KIVQEQKLLA IESGANY--- ----------
OC43                KEINIIKLSG VKKPFKVEDS VIVNDDTSET KYVKSLSIVD VYDMWLTGCK YVVRTANALS
BoCoV               KEINIIKLSG VKKPFKVEDS VIVNDDTSEI KYVKSLSIVD VYDMWLTGCR CVVRTANALS
MHV                 VDVKEVKLNG VKKPIKVEDS VVVNDPTSET KVVKSLSIVD VYDMFLTGCR YVVWMANELS
AIPV                ---------- ---------- ------IAKK AIVGSSVVTT QCGKLIG--- ----------
SARS CoV            PTIQKEVIEC DVKTTEVVGN VILKPSDEGV KVTQELGHED LMAAYVENTS ITIKKPNELS

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2345       2355       2365       2375       2385       2395
EMCR                -------AQK FFQFGDFVMN N--------- ---------- --------IV LFLTWLLSMF
229E                -------AQK FFDFGDFLIH N--------- ---------- --------FV IFFTWLLSMF
PEDV                ------ASER FFSFGDFMSR N--------- ---------- --------LI TVFLYILSIL
TGEV                ------ALTE FGRYADMFFM A--------- ---------- --------GD KILRLLLEVF
OC43                RAVNVPTIRK FIKFGMTLVS IPIDLLNLRE IKPAVNVVKA VRNKISVCFN FIKWLFVLLF
BoCoV               RAVNVPTIRK FIKFGMTLVS IPIDLLNLRE IKPVFNVVKA VRNKISACFN FIKWLFVLLF
MHV                 RLVNSPTVRE YVKWGMTKIV IPAKLVLLRD EKQEFVAPKV VKAKVIACYS AVKWFFLYCF
AIPV                ---------K AATFIADKVG G--------- ---------- ---------G VVRNITDSIK
SARS CoV            LALG---LKT IATHGIAAIN SVPWSKILAY VKPFLG--QA AITTSNCAKR LAQRVFNNYM

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2405       2415       2425       2435       2445       2455
EMCR                SLLRTSIMKH DIKVIAKAPK RTGVILTRSF KYNIRSALFV VKQKWC-VIV TLFKFLLLLY
229E                TLCKTAVTTG DVKIMAKAPQ RTGVVLKRSL KYNLKASAAV LKSKWW-LLA KFTKLLLLIY
PEDV                GLCFRAFRKR DVKVLAGVPQ RTGIILRKSM RYNAKALGVF FKLKLY-WFK VLGKFSLGIY
TGEV                KYLLVLFMCL RSTKMPKVKV KP-PLAFKDF GAKVRTLNYM RQLNKP-SVW RYAKLVLLLI
OC43                GWIKISADNK VIYTTEIASK LTCKLVALAF KNAFLTFKWS MVARGA-CII ATIFLLWFNF
BoCoV               GWIKISADNK VIYTTEVASK LTCKLVALAF KNAFLTFKWS VVARGA-CII ATIFLLWFNF
MHV                 SWIKENTDNK VIYTTEVASK LTFNLCCLAF KNALQTFNWN VVSRGF-FLV ATVFLLWFNF
AIPV                GLCGITRGHF ERKMSPQFLK TLMFFLFYFL KASVKSVVAS YKTVLCKVVL ATLLIVWFVY
SARS CoV            PYVFTLLFQL CTFTKSTNSR IRASLPTTIA KNSVKSVAKL CLDAGI-NYV KSPKFSKLFT

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2465       2475       2485       2495       2505       2515
EMCR                AIYALVFMIV QFSPFNSL-L CGDIVSGYEK STF------- --------NK DIYCGNSMVC
229E                TLYSVVLLCV RFGPFN---F CSETVNGYAK SNF------- --------VK DDYCDGSLGC
PEDV                ALYALLFMTI RFTPIGSP-V CDDVVAGYAN SSF------- --------DK NEYCN-SVIC
TGEV                AIYNFFYLFV SIPVVHKL-T CNGAVQAYKN SSF------- --------IK SAVCGNSILC
OC43                IYANVIFSDF YLPKIGFLPT FVGKIAQWIK NTFSLVTICD LYSMQDVGFK NQYCNGSIAC
BoCoV               IYANVIFSDF YLPKIGFLPT FVGKIVQWIK NTFSLVTICD LYSIQDVGFK NQYCNGSIAC
MHV                 LYANVILSDF YLPNIGFFPT FVGQIVAWVK TTFGIFTLCD LYQVSDVGYR SSFCNGSMVC
AIPV                TSNPVMFTGI RVLDFLFEGS LCGPYKDYGK DSFD------ --------VL R-YCADDFIC
SARS CoV            IAMWLLLLSI CLGSLICVTA AFGVLLSNFG APSYCNGVRE LYLNSSNVTT MDFCEGSFPC

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2525       2535       2545       2555       2565       2575
EMCR                KMCLFSYQEF NDLDHTSLVW KHIR----DP ILISLQPFVI LVILLIFG-- ----------
229E                KMCLFGYQEL SQFSHLDVVW KHIT----DP LFSNMQPFIV MVLLLIFG-- ----------
PEDV                KVCLYGYQEL SDFSHTQVVW QHLR----DP LIGNVMPFFY LAFLAIFG-- ----------
TGEV                KACLASYDEL ADFQHLQVTW DFKS----DP LWNRLVQLSY FAFLAVFG-- ----------
OC43                QFCLAGFDML DNYKAIDVVQ YEAD----RR AFVDYTGVLK IVIELIVSYA LYTAWFYPLF
BoCoV               QFCLAGFDML DNYKAIDVVQ YEAD----RR AFVDYTGVLK IVIELIVSYA LYTAWFYPLF
MHV                 ELCFSGFDML DNYDAINVVQ HVVD----RR VSFDYISLFK LVVELVIGYS LYTVCFYPLF
AIPV                RVCLHDKDSL HLYKHAYSVE QVYKDAASGF IFNWNWLYLV FLILFVKP-- ----------
SARS CoV            SICLSGLDSL DSYPALETIQ VTISSYKLDL TILGLAAEWV LAYMLFTKFF YLLG-----L

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       2585       2595       2605       2615       2625       2635
EMCR                NMYLRFGLLY FVAQFISTFG SFLGFHQKQW FLHFVPFDVL CNEFLATFIV CKIVLFVRHI
229E                DNYLRCFLLY FVAQMISTVG VFLGYKETNW FLHFIPFDVI CDELLVTVIV IKVISFVRHV
PEDV                GVYVKAITLY FIFQYLNSLG VFLGLQQSIW FLQLVPFDVF GDEIVVFFIV TRVLMFIKHV
TGEV                NNYVRCFLMY FVSQYLNLWL SYFGYVEYSW FLHVVNFESI SAEFVIVVIV VKAVLALKHI
OC43                ALISIQILTT WLPELFMLST LHWSFRLLVA LANMLPAHVF MRFYIIIASF IKLFSLFRHV
```

```
                BoCoV       ALISIQILTT WLPELLMLST LHWSVRLLVS LANMLPAHVF MRFYIIIASF IKLFSLFRHV
                MHV         GLIGMQLLTT WLPEFFMLET MHWSARFFVF VANMLPAFTL LRFYIVVTAM YKIFCLCRHV
                AIPV        VAGFVIICYC VKYLVLNSTV LQTGVCFLDW FVQTVFSHFN FMGAGFYFWL FYKIYIQVHH
                SARS CoV    SAIMQVFFGY FASHFISNS- --WLMWFIIS IVQMAPVSAM VRMYIFFASF YYIWKSYVHI
          5
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2645       2655       2665       2675       2685       2695
                EMCR        IVGCNNADCV ACSKSARLKR VPLQTIINGM HKSFYVNANG GTCFCNKHNF FCVNCDSFGP
                229E        LFGCENPDCI ACSKSARLKR FPVNTIVNGV QRSFYVNANG GSKFCKKHRF FCVDCDSYGY
         10     PEDV        CLGCDKASCV ACSKSARLKR VPVQTIFQGT SKSFYVHANG GSKFCKKHNF FCLNCDSYGP
                TGEV        VFACSNPSCK TCSRTARQTR IPIQVVVNGS MKTVYVHANG TGKFCKKHNF YCKNCDSYGF
                OC43        AYGCSKSGCL FCYKRNRSLR VKCSTIVGGM IRYYDVMANG GTGFCSKHQW NCIDCDSYKP
                BoCoV       AYGCSKSGCL FCYKRNRSLR VKCSTIVGGM IRYYDVMANG GTGFCSKHQW NCIDCDSYKP
                MHV         MYGCSRPGCL FCYKRNRSVR VKCSTVVGGT LRYYDVMANG GTGFCAKHQW NCLNCSAFGP
         15     AIPV        ILYCKDVTCE VCKRVARSNR QEVSVVVGGR KQIVHVYTNS GYNFCKRHNW YCRNCDDYGH
                SARS CoV    MDGCTSSTCM MCYKRNRATR VECTTIVNGM KRSFYVYANG GRGFCKTHNW NCLNCDTFCT

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2705       2715       2725       2735       2745       2755
         20     EMCR        GNTFINGDIA RELGNVVKTA VQPTAPAYVI IDKVDFVNGF YRLYSGDTFW RYDFDITESK
                229E        GSTFITPEVS RELGNITKTN VQPTGPAYVM IDKVEFENGF YRLYSCETFW RYNFDITESK
                PEDV        GCTFINDVIA TEVGNVVKLN VQPTGPATIL IDKVEFSNGF YYLYSGDTFW KYNFDITDSK
                TGEV        ENTFICDEIV RDLSNSVKQT VYATDRSHQE VTKVECSDGF YRFYVGDEFT SYDYDVKHKK
                OC43        GNTFITVEAA LDLSKELKRP IQPTDVAYHT VTDVKQVGCS MRLFYDRDGQ RTYDDVNASL
         25     BoCoV       GNTFITVEAA LDLSKELKRP IQPTDVAYHT VTDVKQVGCY MRLFYDRDGQ RTYDDVNASL
                MHV         GNTFITHEAA ADLSKELKRP VNPTDSAYYL VTEVKQVGCS MRLFYERDGQ RVYDDVSASL
                AIPV        QNTFMSPEVA GELSEKLKRH VKPTAYAYHV VDEACLVDDF VNLKYKAATP GKDSASSAVK
                SARS CoV    GSTFISDEVA RDLSLQFKRP INPTDQSSYI VDSVAVKNGA LHLYFDKAGQ KTYERHPLSH

         30                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2765       2775       2785       2795       2805       2815
                EMCR        YSCKEVLKN- ---------- CNVLENFIVY NNSGS--NIT QIKNACVYFS QLLCEPIKLV
                229E        YSCKEVFKN- ---------- CNVLDDFIVF NNNGT--NVT QVKNASVYFS QLLCRPIKLV
                PEDV        YTCKEALKN- ---------- CSIITDFIVF NNNGS--NVN QVKNACVYFS QMLCKPVKLV
         35     TGEV        YSSQEVLKS- ---------- MLLLDDFIVY SPSGS--ALA NVRNACVYFS QLIGKPIKIV
                OC43        FVDYSNLLHS ---------KV KSVPNMHVVV VENDA--DKA NFLNAAVFYA QSLFRPILMV
                BoCoV       FVDYSNLLHS ---------KV KSVPNMHVVV VENDA--DKA NFLNAAVFYA QSLFRPILMV
                MHV         FVDMNGLLHS ---------KV KGVPETHVVV VENEA--DKA GFLNAAVFYA QSLYRPMLLV
                AIPV        CFSVTDFLKK AVFLKEALKC EQISNDGFIV CNTQSAHALE EAKNAAIYYA QYLCKPILIL
         40     SARS CoV    FVNLDNLRAN ---------NT KGSLPINVIV FDGKSKCDES ASKSASVYYS QLMCQPILLL

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2825       2835       2845       2855       2865       2875
                EMCR        NSELLSTLS- -VDFNGVLHK AYVDVLCNSF FKELTANMSM AECKATLGLT ----------
         45     229E        DSELLSTLS- -VDFNGVLHK AYIDVLRNSF GKDLNANMSL AECKRALGLS ----------
                PEDV        DSALLASLS- -VDFGASLHS AFVSVLSNSF GKDLSSCNDM QDCKSTLGFD D----------
                TGEV        NSDLLEDLS- -VDFKGALFN AKKNVIKNSF NVDVSECKNL DECYRACNLN ----------
                OC43        DKNLITTANT GTSVTETMFD VYVDTFLSMF DVDKKSLNAL IATAHSSIKQ GTQIYKVLDT
                BoCoV       DKILITTANT GTSVTETMFD VYVDTFLSMF DVDKKSLNAL IATAHSSIKQ GTQICKVLDT
         50     MHV         EKKLITTANT GLSVSQTMFD LYVDSLLGVL DVDRKSLTSF VNAAHNSLKE GVQLEQVMDT
                AIPV        DQALYEQLVV -EPVSKSVID KVCSILSSII SVDTAALNYK AGTLRDALLS ----------
                SARS CoV    DQVLVSDVGD STEVSVKMFD AYVDTFSATF SVPMEKLKAL VATAHSELAK GVALDGVLST

         55                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2885       2895       2905       2915       2925       2935
                EMCR        ---------- ----VSDDDF VSAVANAHRY DVLLSDLSFN NFFISYAKPE DK-LSVYDIA
                229E        ---------- ----ISDHEF TSAISNAHRC DVLLSDLSFN NFVSSYAKPE EK-LSAYDLA
                PEDV        ---------- ----VPLDTF NAAVAEAHRY DVLLTDMSFN NFTTSYAKPE EK-FPVHDIA
                TGEV        ---------- ----VSFSTF EMAVNNAHRF GILITDRSFN NFWPSKVKPG SSGVSAMDIG
         60     OC43        FLSCARKSCS IDSDVDTKCL ADSVMSAVSA GLELTDESCN NLVPTYLKSD N--IVAADLG
                BoCoV       FLSCARKSCS IDSDVDTKCL ADSVMSAVSA GLELTDESCN NLVPTYLKGD N--IVAADLG
                MHV         FIGCARRKCA IDSDVETKSI TKSIMSAVNA GVDFTDESCN NLVPTYVKSD T--IVAADLG
                AIPV        ---------- ---ITKDEEA VDMAIFCHNH DVDYTGDGFT NVIPSYGIDT G-KLTPRDRG
                SARS CoV    FVSAARQG-V VDTDVDTKDV IECLKLSHHS DLEVTGDSCN NFMLTYNKVE N--MTPRDLG
         65
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2945       2955       2965       2975       2985       2995
                EMCR        CCMRAGSKVV NHNVLIKESI PIVWGVKDFN TLSQEGKKYL VKTTKAKGLT FLLTFNDNQA
                229E        CCMRAGAKVV NANVLTKDQT PIVWHAKDFN SLSAEGRKYI VKTSKAKGLT FLLTINENQA
         70     PEDV        TCMRVGAKIV NHNVLVKDSI PVVWLVRDFI ALSEETRKYI IRTTKVKGIT FMLTFNDCRM
                TGEV        KCMTSDAKIV NAKVLTQRGK SVVWLSQDFA ALSSTAQKVL VKTFVEEGVN FSLTFNAVGS
                OC43        VLIQNSAKHV QGNVAKIAGV SCIWSVDAFN QFSSDFQHKL KKACCKTGLK LKLTYNKQMA
                BoCoV       VLIQNSAKHV QGNVAKIAGV SCIWSVDAFN QLSSDFQHKL KKACCKTGLK LELTYNKQMA
                MHV         VLIQNNAKHV QANVAKAANV ACIWSVDAFN QLSADLQHRL RKACSKTGLK IKLTYNKQEA
         75     AIPV        FLINADASIA NLRVKN--AP PVVWKFSELI KLSDSCLKYL ISATVKSGVR FFITKSGAKQ
                SARS CoV    ACIDCNARHI NAQVAKSHNV SLIWNVKDYM SLSEQLRKQI RSAAKKNNIP FRLTCATTRQ

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3005       3015       3025       3035       3045       3055
         80     EMCR        ITQVP----A TSIVAKQGAG ---------- FKRTYNFLWY VCLFVVALFI GVSFID----
```

```
                229E        VTQIP----A TSIVAKQGAG D--------- AGHSLTWLWL LCGLVCLIQF YLCFFMPY--
                PEDV        HTTIP----T VCIANKKGAG LP--------S FSKVKKFFWF LCLFIVAAFF ALSFLD----
                TGEV        DDDLPYERFT ESVSPKSGSG ---------- FFDVITQLKQ IVILVFVFIF ICGLCSVYSV
                OC43        NVSVL----T TPFSLKGGAV FS--------Y FVYVCFVLSL VCFIGLWCLM PTYTVH----
                BoCoV       NVSVL----T TPFSLKGGAV FS--------Y FVYVCFVLSL VCFIGLWCLM PTYTVH----
                MHV         NVPIL----T TPFSLKGGAV FS--------K VLQWLFVVNL ICFIVLWALM PTYAVH----
                AIPV        VIACHT--QK LLVEKKAGGI VSGTFKCFKS YFKWLLIFYI LFTACCSGYY YMEVSKSFVH
                SARS CoV    VVNVI----T TKISLKGGKI VS--------T CFKLMLKATL LCVLAALVCY IVMPVHTLS-

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3065       3075       3085       3095       3105       3115
                EMCR        -YTTTVTSFH GYDFKYIENG QLKVFEAPLH CVRNVFDNFN QWHEAKFGVV TTNSD-KCPI
                229E        FMYDIVSSFE GYDFKYIENG QLKNFEAPLK CVRNVFENFE DWHYAKFGFT PLNKQ-SCPI
                PEDV        -FSTQVSSDS DYDFKYIESG QLKTFDNPLS CVHNVFINFD QWHDAKFGFT PVNNP-SCPI
                TGEV        ATQSYIESAE GYDYMVIKNG IVQPFDDTIS CVHNTYKGFG DWFKAKYGFI PTFGK-SCPI
                OC43        --KSDFQLPV YASYKVLDNG VIRDVSVEDV CFANKFEQFD QWYESTFGLS YYSNSMACPI
                BoCoV       --KSDFQLPV YASYKVLDNG VIRDVSVEDV CFANKFEQFD QWYESTFGLS YYSNSMACPI
                MHV         --KSDMQLPL YASFKVIDNG VLRDVTVTDA CFANKFIQFD QWYESTFGLV YYRNSRACPV
                AIPV        PMYDVNSTLH VEGFKVIDKG VLREIVPEDT CFSNKFVNFD AFWG----RP YDNSR-NCPI
                SARS CoV    --IHDGYTNE IIGYKAIQDG VTRDIISTDD CFANKHAGFD AWFSQRGGSY KNDKS--CPV

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3125       3135       3145       3155       3165       3175
                EMCR        VVG---VSER INVVPGVPTN VYLVG------ ------KTLV FTLQAAFGNT GVCYDFDGVT
                229E        VVG---VSEI VNTVAGIPSN VYLVG------ ------KTLI FTLQAAFGNA GVCYDIFGVT
                PEDV        VVG---VSDE ARTVPGIPAG VYLAG------ ------KTLV FAINTIFGTS GLCFDASGVA
                TGEV        VVGTVFDLEN MRPIPDVPAY VSIVG------ ------RSLV FAINAAFGVT NMCYDHTGNA
                OC43        VVA-VIDQDF GSTVFNVPTK VLRYG------ ------YHVL HFITHALSAD GVQCYTPHSQ
                BoCoV       VVA-VVDQDF GSTVFNVPTK VLRYG------ ------YHVL HFITHALSAD GVQCYTPHSQ
                MHV         VVA-VIDQDI GYTLFNVPTK VLRYG------ ------FHVL HFITHAFATD SVQCYTPHMQ
                AIPV        VTA--VIDGD GTVATGVPGF VSWVMDGVMF IHMTQTERKP WYIPTWFNRE IVGYTQDSII
                SARS CoV    VAA-IITREI GFIVPGLPGT VLRAIN----- ------GDFL HFLPRVFSAV GNICYTPSKL

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3185       3195       3205       3215       3225       3235
                EMCR        TS------DK CIFNSACTRL EGLGGD-NVY CYN-TDLIEG SKPYSILQPN AYYKYDVKN-
                229E        TP------EK CIFTSACTRL EGLGGN-NVY CYN-TALMEG SLPYSSIQAN AYYKYDNGN-
                PEDV        DK------GA CIFNSACTTL SGLGGT-AVY CYK-NGLVEG AKLYSELAPH SYYKMVDGN-
                TGEV        VSKD-SYFDT CVFNTACTTL TGLGGT-IVY CAK-QGLVEG AKLYSDLMPD YYYEHASGN-
                OC43        ISYSNFYASG CVLSSACTMF TMADGSPQPY CYT-EGLMQN ASLYSSLVPH VRYNLANAKG
                BoCoV       ISYSNFYASG CVLSSACTMF AMADGSPQPY CYT-DGLMQN ASLYSSLVPH VRYNLANAKG
                MHV         IPYDNFYASG CVLSSLCTML AHADGTPHPY CYT-EGIMHN ASLYDSLAPH VRYNLANSNG
                AIPV        TEG-SFYTSI ALFSARCLYL TASNTP-QLY CFNGDNDAPG ALPFGSIIPH RVYFQPNGVR
                SARS CoV    IEYSDFATSA CVLAAECTIF KDAMGKPVPY CYD-TNLLEG SISYSELRPD TRYVLMDGS-

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3245       3255       3265       3275       3285       3295
                EMCR        YVRFPEILAR GFGLRTIRTL ATRYCRVGEC RDSHKGVCFG FDKWYVNDGR VD---DGYIC
                229E        FIKLPEVIAQ GFGFRTVRTI ATKYCRVGEC VESNAGVCFG FDKWFVNDGR VA---NGYVC
                PEDV        AVSLPEIISR GFGIRTIRTK AMTYCRVGQC VQSAEGVCFG ADRFFVYNAE SG---SDFVC
                TGEV        MVKLPAIIR- GLGLRFVKTQ ATTYCRVGEC IDSKAGFCFG GDNWFVYDNE FG---NGYIC
                OC43        FIRFPEVLRE GL-VRIVRTR SMSYCRVGLC EEADEGICFN FNGSWVLNND YYRSLPGTFC
                BoCoV       FIRLPEVLRE GL-VRIVRTR SMSYCRVGLC EEADEGICFN FNGSWVLNND YYRSLPGTFC
                MHV         YIRFPEVVSE GI-VRIVRTR SMTYCRVGLC EDAEEGVCFN FNSSWVLNNP YYRAMPGTFC
                AIPV        LIVPQQILHT PY---VVKFV SDSYCRGSVC EYTRPGYCVS LNPQWVLFND EYTSKPGVFC
                SARS CoV    IIQFPNTYLE GS-VRVVTTF DAEYCRHGTC ERSEVGICLS TSGRWVLNNE HYRALSGVFC

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3305       3315       3325       3335       3345       3355
                EMCR        GDGLIDLLVN VLSIFSSSFS VVAMSGHMLF NFLFAAFITF LCFLVTKFKR VFGDLSYGVF
                229E        GTGLWNLVFN ILSMFSSSFS VAAMSGQILL NCALGAFAIF CCFLVTKFRR MFGDLSVGVC
                PEDV        GTGLFTLLMN VISVFSKTVP VTVLSGQILF NCIIAFVAVA VCFLFTKFKR MFGDMSVGVF
                TGEV        GNSVLGFFKN VFKLFNSNMS VVATSGAMLV NIIIACLAIA MCYGVLKFKK IFGDCTFLIV
                OC43        GRDVFDLIYQ LFKGLAQPVD FLALTASSIA GAILAVIVVL VFYYLIKLKR AFGDYTSVVF
                BoCoV       GRDVFDLIYQ LFKGLAQPVD FLALTASSIA GAILAVIVVL GFYYLIKLKR AFGDYTSIVF
                MHV         GRNAFDLIHQ VLGGLVRPID FFALTASSVA GAILAIIVVL AFYYLIKLKR AFGDYTSIVF
                AIPV        GSTVRELMFS MVSTFFTGVN -PNIYMQLAT MFLILVVVVL IFAMVIKFQG VFKAYATTVF
                SARS CoV    GVDAMNLIAN IFTPLVQPVG ALDVSASVVA GGIIAILVTC AAYYFMKFRR VFGEYNHVVA

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              3365       3375       3385       3395       3405       3415
                EMCR        TVVCATLINN ISYVVTQN-L FFMLLYAILY FVFTRTVR-- YAWIWHIAYI VAYFLLIPWW
                229E        TVVVAVLLNN VSYIVTQN-L VTMIAYAILY FFATRSLR-- YAWIWCAAYL IAYISFAPWW
                PEDV        TVGACTLLNN VSYIVTQN-T LGMLGYATLY FLCTKGVR-- YMWIWHLGFL ISYILIAPWW
                TGEV        MIIVTLVVNN VSYFVTQN-T FFMIIYAIVY YFITRKLA-- YPGILDAGFI IAYINMAPWY
                OC43        VNVIVWCVNF MMLFVFQVYP ILSCVYAICY FYATLYFPSE ISVIMHLQWL VMYGTIMPLW
                BoCoV       VNVIVWCVNF MMLFVFQVYP TLSCVYAICY FYATLYFPSE ISVIMHLQWL VMYGTIMPLW
                MHV         INVIVWCINF LMLFVFQVYP TLSCLYACFY FYTTLYFPSE ISVVMHLQWL VMYGAIMPLW
                AIPV        ITMLVWVINA FILCVHSYNS VLAVILLVLY CYASLVTSRN TVIIMHCWLV FTFGLIVPTW
                SARS CoV    ANALLFLMSF TILCLVPAYS FLPGVYSVFY LYLTFYFTND VSFLAHLQWF AMFSPIVPFW
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3425       3435       3445       3455       3465       3475
EMCR                LLTWFSFAAF LELLPNVFKL K----ISTQL FEGDKFIGTF ESAAAGTFVL DMRSYERLIN
229E                LCAWYFLAML TGLLPSLLKL K----VSTNL FEGDKFVGTF ESAAAGTFVI DMRSYEKLAN
PEDV                VLMVYAFSAI FEFMPNLFKL K----VSTQL FEGDKFVGSF ENAAAGTFVL DMHAYERLAN
TGEV                VITAYILVFL YDSLPSLFKL K----VSTNL FEGDKFVGNF ESAAMGTFVI DMRSYETIVN
OC43                FCLLYIAVVV SN--HAFWVF S----YCRKL GTSVRSDGTF EEMALTTFMI TKDSYCKLKN
BoCoV               FCLLYISVVV SN--HAFWVF S----YCRQL GTSVRSDGTF EEMALTTFMI TKDSYCKLKN
MHV                 FCIIYVAVVV SN--HALWLF S----YCRKL GTEVRSDGTF EEMSLTTFMI TKESYCKLKN
AIPV                LACCYLGFII YMYTPLFLWC YGTTKNTRKL YDGNEFVGNY DLAAKSTFVI RGSEFVKLTN
SARS CoV            ITAIYVFCIS LKHCHWFFNN Y----LRKRV MFNGVTFSTF EEAALCTFLL NKEMYLKLRS

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3485       3495       3505       3515       3525       3535
EMCR                T--ISPEKLK NYAASYNKYK YYSGSASEAD YRCACYAHLA KAMLDYAKDH N-DMLYSPPT
229E                S--ISPEKLK SYAASYNRYK YYSGNANEAD YRCACYAYLA KAMLDFSRDH N-DILYTPPT
PEDV                S--ISTEKLR QYASTYNKYK YYSGSASEAD YRLACFAHLA KAMMDYASNH N-DTLYTPPT
TGEV                S--TSIARIK SYANSFNKYK YYTGSMGEAD YRMACYAHLG KALMDYSVNR T-DMLYTPPT
OC43                S--LSDVAFN RYLSLYNKYR YYSGKMDTAA YREAACSQLA KAMDTFTNNN GSDVLYQPPT
BoCoV               S--LSDVAFN RYLSLYNKYR YYSGKMDTAA YREAACSQLA KAMDTFTNNN GSDVLYQPPT
MHV                 S--VSDVAFN RYLSLYNKYR YFSGKMDTAA YREAACSQLA KAMETFNHNN GNDVLYQPPT
AIPV                E--I-GDKFE AYLSAYARLK YYSGTGSEQD YLQACRAWLA YALDQYR-NS GVEIVYTPPR
SARS CoV            ETLLPLTQYN RYLALYNKYK YFSGALDTTS YREAACCHLA KALNDFS-NS GADVLYQPPQ

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3545       3555       3565       3575       3585       3595
EMCR                ISYN-STLQS GLKKMAQPSG CVERCVVRVC YGSTVLNGVW LGDTVTCPRH VIAPS-TTVL
229E                VSYG-STLQA GLRKMAQPSG FVEKCVVRVC YGNTVLNGLW LGDIVYCPRH VIASN-TTSA
PEDV                VSYN-STLQA GLRKMAQPSG VVEKCIVRVC YGNMALNGLW LGDIVMCPRH VIASS-TTST
TGEV                VSVN-STLQS GLRKMAQPSG LVEPCIVRVS YGNNVLNGLW LGDEVICPRH VIASD-TTRV
OC43                ASVSTSFLQS GIVKMVNPTS KVEPCVVSVT YGNMTLNGLW LDDKVYCPRH VICSASDMTN
BoCoV               ASVSTSFLQS GIVKMVNPTS KVEPCIVSVT YGNMTLNGLW LDDKVYCPRH VICSASDMTN
MHV                 ASVTTSFLQS GIVKMVFPTS KVEPCVVSVT YGNMTLNGLW LDDKVYCPRH VICSSADMTD
AIPV                YSIGVSRLQS GFKKLVSPSS AVEKCIVSVS YRGNNLNGLW LGDTIYCPRH VLG---KFSG
SARS CoV            TSITSAVLQS GFRKMAFPSG KVEGCMVQVT CGTTTLNGLW LDDTVYCPRH VICTAEDMLN

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3605       3615       3625       3635       3645       3655
EMCR                IDYDHAYSTM RLHNFSVSHN G-VFLGVVGV TMHGSVLRIK VSQSNVHTPK HVFKTLKPGA
229E                IDYDHEYSIM RLHNFSIISG T-AFLGVVGA TMHGVTLKIK VSQTNMHTPR HSFRTLKSGE
PEDV                IDYDYALSVL RLHNFSISSG N-VFLGVVSA TMRGALLQIK VNQNNVHTPK YTYRTVRPGE
TGEV                INYENEMSSV RLHNFSVSKN N-VFLGVVSA RYKGVNLVLK VNQVNPNTPE HKFKSIKAGE
OC43                PDYTNLLCRV TSSDFTVLFD R-LSLTVMSY QMRGCMLVLT VTLQNSRTPK YTFGVVKPGE
BoCoV               PDYTNLLCRV TSSDFTVLFD R-LSLTVMSY QMQGCMLVLT VTLQNSRTPK YTFGVVKPGE
MHV                 PDYSNLLCRV ISSDFCVMSG R-MSLTVMSY QMQGSLLVLT VTLQNPNTPK YSFGVVKPGE
AIPV                DQWNDVLNLA NNHEFEVTTQ HGVTLNVVSR RLKGAVLILQ TAVANAETPK YKFIKANCGD
SARS CoV            PNYEDLLIRK SNHSFLVQAG N-VQLRVIGH SMQNCLLRLK VDTSNPKTPK YKFVRIQPGQ

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3665       3675       3685       3695       3705       3715
EMCR                SFNILACYEG IASGVFGVNL RTNFTIKGSF INGACGSPGY NVRNDGTVEF CYLHQIELGS
229E                GFNILACYDG CAQGVFGVNM RTNWTIRGSF INGACGSPGY NLKN-GEVEF VYMHQIELGS
PEDV                SFNILACYDG AAAGVYGVNM RSNYTIRGSF INGACGSPGY NINN-GTVEF CYLHQLELGS
TGEV                SFNILACYEG CPGSVYGVNM RSQGTIKGSF IAGTCGSVGY VLEN-GILYF VYMHHLELGN
OC43                TFTVLAAYNG KPQGAFHVTM RSSYTIKGSF LCGSCGSVGY VIMG-DCVKF VYMHQLELST
BoCoV               TFTVLAAYNG KPQGAFHVTM RSSYTIKGSF LCGSCGSVGY VIMG-DCVKF VYMHQLELST
MHV                 TFTVLAAYNG KSQGAFHVTM RSSYTIKGSF LCGSCGSVGY VLTG-DSVRF VYMHQLELST
AIPV                SFTIACAYGG TVVGLYPVTM RSNGTIRASF LAGACGSVGF NIEK-GVVNF FYMHHLELPN
SARS CoV            TFSVLACYNG SPSGVYQCAM RPNHTIKGSF LNGSCGSVGF NIDY-DCVSF CYMHHMELPT

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3725       3735       3745       3755       3765       3775
EMCR                GAHVGSDFTG SVYGNFDDQP SLQVESANLM LSDNVVAFLY AALLNGCR-- ----WWLRST
229E                GSHVGSSFDG VMYGGFEDQP NLQVESANQM LTVNVVAFLY AAILNGCT-- ----WWLKGE
PEDV                GCHVGSDLDG VMYGGYEDQP TLQVEGASSL FTENVLAFLY AALINGST-- ----WWLSSS
TGEV                GSHVGSNFEG EMYGGYEDQP SMQLEGTNVM SSDNVVAFLY AALINGER-- ----WFVTNT
OC43                GCHTGTDFNG DFYGPYKDAQ VVQLLIQDYI QSVNFVAWLY AAILNNCN-- ----WFVQSD
BoCoV               GCHTGTDFNG DFYGPYKDAQ VVQLPVQDYI QSVNFVAWLY AAILNNCN-- ----WFVQSD
MHV                 GCHTGTDFSG NFYGPYRDAQ VVQLPVQDYT QTVNVVAWLY AAILNRCN-- ----WFVQSD
AIPV                ALHTGTDLMG EFYGGYVDEE VAQRVPPDNL VTNNIVAWLY AAIISVKESS FSLPKWLEST
SARS CoV            GVHAGTDLEG KFYGPFVDRQ TAQAAGTDTT ITLNVLAWLY AAVINGDR-- ----WFLNRF

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      3785       3795       3805       3815       3825       3835
EMCR                RVNVDGFNEW AMANGYTIVS SV--ECYSIL AAKTGVSVEQ LLASIQHLHE -GFGGKNILG
229E                KLFVEHYNEW AQANGFTAMN GE--DAFSIL AAKTGVCVER LLHAIQVLNN -GFGGKQILG
PEDV                RIAVDRFNEW AVHNGMTTVG NT--DCFSIL AAKTGVDVQR LLASIQSLHK -NFGGKQILG
TGEV                SMSLESYNTW AKTNSFTELS ST--DAFSML AAKTGQSVEK LLDSIVRLNK -GFGGRTILS
OC43                KCSVEDFNVW ALSNGFSQVK SD--LVIDAL ASMTGVSLET LLAAIKRLKN -GFQGRQIMG
```

```
                BoCoV       KCSVEDFNVW ALSNGFSQVK SD--LVIDAL ASMTGVSLET LLAAIKRLKN -GFQGRQIMG
                MHV         SCSLEEFNVW AMTNGFSSIK AD--LVLDAL ASMTGVTVEQ ILAAIKRLYS -GFQGKQILG
                AIPV        TVSVDDYNKW AGDNGFTPFS TS--TAITKL SAITGVDVCK LLRTIMVKNS -QWGGDPILG
                SARS CoV    TTTLNDFNLV AMKYNYEPLT QDHVDILGPL SAQTGIAVLD MCAALKELLQ NGMNGRTILG
5
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3845       3855       3865       3875       3885       3895
                EMCR        YSSLCDEFTL AEVVKQMYGV NLQSGK---V IFGLKTMFLF SVFFTMFWAE LFIYTNTIWI
                229E        YSSLNDEFSI NEVVKQMFGV NLQSGK---T TSMFKSISLF AGFFVMFWAE LFVYTTTIWV
10              PEDV        HTSLTDEFTT GEVVRQMYGV NLQGGY---V SRACRNVLLV GSFLTFFWSE LVSYTKFFWV
                TGEV        YGSLCDEFTP TEVIRQMYGV NLQAGK---V KSFFYPIMTA MTILFAFWLE FFMYTPFTWI
                OC43        SCSFEDELTP SDVYQQLAGI KLQSKRTRLF KGTVCWIMAS TFLFSCIITA FVKWTMFMYV
                BoCoV       SCSFEDELTP SDVYQQLAGI KLQSKRTRLV KGIVCWIMAS TFLFSCIITA FVKWTMFMYV
                MHV         SCVLEDELTP SDVYQQLAGV KLQSKRTRVV KGTCCWILAS TLLFCSIISA FVKWTMFMYV
15              AIPV        QYNFEDELTP ESVFNQIGGV RLQSSFVR-- K--ATSWFWS RCVLACFLFV LCAIVLFTAV
                SARS CoV    STILEDEFTP FDVVRQCSGV TFQGKFKKIV KGTHHWMLLT FLTSLLILVQ STQWSLFFFV

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3905       3915       3925       3935       3945       3955
20              EMCR        NPVILTPIFC LLLFLSLVLT MFLKHKFLFL QVFLLPTVIA TALYNC-VLD YYIVKFLADH
                229E        NPGFLTPFMI LLVALSLCLT FVVKHKVLFL QVFLLPSIIV AAIQNC-AWD YHVTKVLAEK
                PEDV        NPGYVTPMFA CLSLLSSLLM FTLKHKTLFF QVFLIPALIV TSCINL-AFD VEVYNYLAEH
                TGEV        NPTFVSIVLA VTTLISTVFV SGIKHKMLFF MSFVLPSVIL VTAHNL-FWD FSYYESLQSI
                OC43        TTNMFSITFC ALCVIS-LAM LLVKHKHLYL TMYITP-VLF TLLYNN-YLV VYKHTFRGYV
25              BoCoV       TTNMLSITFC ALCVIS-LAM LLVKHKHLYL TMYIIP-VLF TLLYNN-YLV VYKQTFRGYV
                MHV         TTHMLGVTLC ALCFVS-FAM LLVKHKHLYL TMFIMP-VLC TLFYTN-YLV VYKQSFRGLA
                AIPV        PLKFYVYAAV ILLMAVLFIS FTVKHVMAYM DTFLLPTLIT VIIGVCAEVP FIYNTLISQV
                SARS CoV    YENAFLPFTL GIMAIAACAM LLVKHKHAFL CLFLLPSLAT VAYFN---MV YMPASWVMRI
30
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3965       3975       3985       3995       4005       4015
                EMCR        FN-YNVSVLQ MDVQGLVNVL VCLFVVFLH- --TWRFSKER FTHWFTYVCS LIAVAYTYFY
                229E        FD-YNVSVMQ MDIQGFVNIF ICLFVALLH- --TWRFAKER CTHWCTYLFS LIAVLYTALY
                PEDV        FD-YHVSLMG FNAQGLVNIF VCFVVTILHG TYTWRFFN-T PASSVTYVVA LLTAAYNYFY
35              TGEV        VENTNTMFLP VDMQGVMLTV FCFIVFVTYS VRFFTCKQSW FSLAVTTILV IFNMVKIFGT
                OC43        YAWLSYYVPS VEYTYTDEVI YGMLLLVGMV FVTLRSINHD LFSFIMFVGR LISVFSLWYK
                BoCoV       YAWLSYYVPS VEYTYTDEVI YGMLLLIGMV FVTLRSINHD LFSFIMFVGR VISVVSLWYM
                MHV         YAWLSHFVPA VDYTYMDEVL YGVVLLVAMV FVTMRSINHD VFSVMFLVGR LVSLVSMWYF
                AIPV        VIFLSQWYDP VVFDTMVPWM FLPLVLYTAF KCVQGCYMNS FNTSLLMLYQ FVKLGFVIYT
40              SARS CoV    MTWLELADTS LSGYRLKDCV MYASALVLLI LMTARTVYDD AARRVWTLMN VITLVYKVYY

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  4025       4035       4045       4055       4065       4075
                EMCR        SGD------- ------FLSL LVMFLCAISS DWYIGAIVFR LSRLIIFFSP E---SVFSVF
45              229E        SYD------- ------YVSL LVMLLCAISN EWYIGAIIFR ICRFGVAFLP V---EYVSYF
                PEDV        ASD------- ------ILSC AMTLFASVTG NWFVGAVCYK VAVYMALRFP ----TFVAIF
                TGEV        SDEPWTENQI ---AFCFVNM LTMIVSLTTK DWMVVIASYR IAYYIVVCVM P-S-AFVSDF
                OC43        GSN------- -----LEEEI LLMLASLFGT YTWTTVLSMA VAKVIAKWVA VNV-LYFTDI
                BoCoV       GSN------- -----LEEEI LLMLASLFGT YTWTTALSMA AAKVIAKWVA VNV-LYFTDI
50              MHV         GAN------- -----LEEEV LLFLTSLFGT YTWTTMLSLA TAKVIAKWLA VNV-LYFTDV
                AIPV        SSNTLTAYTE GNWELFFELV HTTVLANVSS NSLIGLFVFK CAKWMLYYCN ------ATYL
                SARS CoV    GNALD----- -----QAISM WALVISVTSN YSGVVTTIMF LARAIVFVCV EYYPLLFITG

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
55                                4085       4095       4105       4115       4125       4135
                EMCR        GDVKLTLVVY LICGYLVCTY WGILYWFNRF FKCTMGVYDF KVSAAEFKYM VANGLHAPYG
                229E        DGVKTVLLFY MLLGFVSCMY YGLLYWINRF CKCTLGVYDF CVSPAEFKYM VANGLNAPNG
                PEDV        GDIKSVMFCY LVLGYFTCCF YGILYWFNRF FKVSVGVYDY TVSAAEFKYM VANGLRAPTG
                TGEV        GFMKCISIVY MACGYLFCCY YGILYWVNRF TCMTCGVYQF TVSAAELKYM TANNLSAPKN
60              OC43        PQIKIVLLCY LFIGYIISCY WGLFSLMNSL FRMPLGVYNY KISVQELRYM NANGLRPPKN
                BoCoV       PQIKIVLVCY LFIGYIISCY WGLFSLMNSL FRMPLGVYNY KISVQELRYM NANGLRPPKN
                MHV         PQVKLVLLSY LCIGYVCCCY WGVLSLLNSI FRMPLGVYNY KISVQELRYM NANGLRPPRN
                AIPV        NNYVLMAVMV NCIGWLCTCY FGLYWWVNKV FGLTLGKYNF KVSVDQYRYM CLHKINPPKT
                SARS CoV    NTLQCIMLVY CFLGYCCCCY FGLFCLLNRY FRLTLGVYDY LVSTQEFRYM NSQGLLPPKS
65
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  4145       4155       4165       4175       4185       4195
                EMCR        PFDALWLSFK LLGIGGDRCI KISTVQSKLT DLKCTNVVLL GCLSSMNIAA NSSEWAYCVD
                229E        PFDALFLSFK LMGIGGPRTI KVSTVQSKLT DLKCTNVVLM GILSNMNIAS NSKEWAYCVE
70              PEDV        TLDSLLLSAK LIGIGGERNI KISSVQSKLT DIKCSNVVLL GCLSSMNVSA NSTEWAYCVD
                TGEV        AYDAMILSAK LIGVGGKRNI KISTVQSKLT EMKCTNVVLL GLLSKMHVES NSKEWNYCVG
                OC43        SFEALMLNFK LLGIGGVPII EVSQFQSKLT DVKCANVVLL NCLQHLHVAS NSKLWHYCST
                BoCoV       SFEALMLNFK LLGIGGVPII EVSQFQSKLT DVKCANGGLL NCLQHLHVAS NSKLWQYCST
                MHV         SFEALVLNFK LLGIGGVPVI EVSQIQSRLT DVKCVNVVLL NCLQHLHIAS SSKLWQYCST
75              AIPV        VWEVFSTNIL IQGIGGDRVL PIATVQAKLS DVKCTTVVLM QLLTKLNVEA NSKMHVYLVE
                SARS CoV    SIDAFKLNIK LLGIGGKPCI KVATVQSKMS DVKCTSVVLL SVLQQLRVES SSKLWAQCVQ

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  4205       4215       4225       4235       4245       4255
80              EMCR        LHNKINLCDD PEKAQGMLLA LLAFFLSKHS DFG------L DGLIDSYFDN SSTLQSVASS
```

```
                229E        MHNKINLCDD PETAQELLLA LLAFFLSKHS DFG------L GDLVDSYFEN DSILQSVASS
                PEDV        LHNKINLCND PEKAQEMLLA LLAFFLSKNS AFG------L DDLLESYFND NSMLQSVAST
                TGEV        LHNEINLCDD PEIVLEKLLA LIAFFLSKHN TCD------L SELIESYFEN TTILQSVASA
                OC43        LHNEILATSD LSVAFEKLAQ LLIVLFANPA AVDSKCLTSI EEVCDDYAKD NTVLQALQSE
                BoCoV       LHNEILATSD LGVAFEKLAQ LLIVLFANPA AVDSKCLTSI EEVCDDYAKD NTVLQALQSE
                MHV         LHNEILATSD LSVAFDKLAQ LLVVLFANPA AVDSKCLASI EEVSDDYVRD STVLQALQSE
                AIPV        LHNKILASDD VGECMDNLLG MLITLFCIDS TID------L SEYCDDILKR STVLQSVTQE
                SARS CoV    LHNDILLAKD TTEAFEKMVS LLSVLLSMQG AVD------I NRLCEEMLDN RATLQAIASE

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4265       4275       4285       4295       4305       4315
                EMCR        FVSMPSYIAY ENARQAYEDA IANGSS---- SQLIKQLKRA MNIAKSEFDH EISVQKKINR
                229E        FVGMPSFVAY ETARQEYENA VANGSS---- PQIIKQLKKA MNVAKAEFDR ESSVQKKINR
                PEDV        YVGLPSYVIY ENARQQYEDA VNNGSP---- PQLVKQLRHA MNVAKSEFDR EASTQRKLDR
                TGEV        YAALPSWIAL EKARADLEEA KKNDVS---- PQILKQLTKA FNIAKSDFER EASVQKKLDK
                OC43        FVNMASFVEY EVAKKNLDEA RFSGSAN--- QQQLKQLEKA CNIAKSAYER DRAVAKKLER
                BoCoV       FVNMASFVEY EVAKKNLDEA CSSGSAN--- QQQLKQLEKA CNIAKSAYER DRAVARKLER
                MHV         FVNMASFVEY ELAKKNLDEA KASGSAN--- QQQIKQLEKA CNIAKSAYER DRAVARKLER
                AIPV        FSHIPSYAEY ERAKNLYEKV LVDSKNGGVT QQELAAYRKA ANIAKSVFDR DLAVQKKLDS
                SARS CoV    FSSLPSYAAY ATAQEAYEQA VANGDS---- EVVLKKLKKS LNVAKSEFDR DAAMQRKLEK

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4325       4335       4345       4355       4365       4375
                EMCR        MAEQAATQMY KEARSVNRKS KVISAMHSLL FGMLRRLDMS SVETVLNLAR DGVVPLSVIP
                229E        MAEQAAAAMY KEARAVNRKS KVVSAMHSLL FGMLRRLDMS SVDTILNMAR NGVVPLSVIP
                PEDV        MAEQAAAQMY KEARAVNRKS KVVSAMHSLL FGMLRRLDMS SVDTILNLAK DGVVPLSVIP
                TGEV        MAEQAAASMY KEARAVDRKS KIVSAMHSLL FGMLKKLDMS SVNTIIDQAR NGVLPLSIIP
                OC43        MADLALTNMY KEARINDKKS KVVSALQTML FSMVRKLDNQ ALNSILDNAV KGCVPLNAIP
                BoCoV       MADLALTNMY KEARINDKKS KVVSALQTML FSMVRKLDNQ ALNSILDNAV KGCVPLNAIP
                MHV         MADLALTNMY KEARINDKKS KVVSALQTML FSMIRKLDNQ ALNSILDNAV KGCVPLNAIP
                AIPV        MAERAMTTMY KEARVTDRRA KLVSSLHALL FSMLKKIDSE KLNVLFDQAS SGVVPLATVP
                SARS CoV    MADQAMTQMY KQARSEDKRA KVTSAMQTML FTMLRKLDND ALNNIINNAR DGCVPLNIIP

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4385       4395       4405       4415       4425       4435
                EMCR        ATSASKLTIV SPDLESYSKI VCDGSVHYAG VVWTLNDVKD NDGRPVHVKE ITR-----EN
                229E        ATSAARLVVV VPDHDSFVKM MVDGFVHYAG VVWTLQEVKD NDGKNVHLKD VTK-----EN
                PEDV        AVSATKLNIV TSDIDSYNRI QREGCVHYAG TIWNIIDIKD NDGKVVHVKE VTA-----QN
                TGEV        AASATRLVVI TPSLEVFSKI RQENNVHYAG AIWTIVEVKD ANGSHVHLKE VTA-----AN
                OC43        SLAANTLNII VPDKSVYDQV VDNVYVTYAG NVWQIQTIQD SDGTNKQLNE IS---------
                BoCoV       SLAANTLTII VPDKSVYDQV VDNVYVTYAG NVWQIQTIQD SDGTNKQLHE IS---------
                MHV         SLTSNTLTII VPDKQVFDQV VDNVYVTYAG NVWHIQSIQD ADGAVKQLNE ID---------
                AIPV        IVCSNKLTLV IPDPETWVKC VEGVHVTYST VVWNIDTVID ADGTELHPTS TGSGLTYCIS
                SARS CoV    LTTAAKLMVV VPDYGTYKNT CDGNTFTYAS ALWEIQQVVD ADSKIVQLSE INM-----DN

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4445       4455       4465       4475       4485       4495
                EMCR        VETLTWPLIL NCER------ VVKLQNNEIM PGKLKQKPMK AEG--DGGVL GDGNALYNTE
                229E        QEILVWPLIL TCER------ VVKLQNNEIM PGKMKVKATK GEG--DGGIT SEGNALYNNE
                PEDV        AESLSWPLVL GCER------ IVKLQNNEII PGKLKQRSIK AEG--DG-IV GEGKALYNNE
                TGEV        ELNLTWPLSI TCER------ TTKLQNNEIM PGKLKERAVR ASATLDGEAF GSGKALMASE
                OC43        -DDCNWPLVI IANRY-NEVS ATVLQNNELM PAKLKIQVVN SGP--DQTCN TPTQCYYNNS
                BoCoV       -DDCNWPLVI IANRH-NEVS ATVLQNNELM PAKLKTQVVN SGP--DQTCN TPTQCYYNNS
                MHV         -VNITWPLVI AANRH-NEVS SVVLQNNELM PQKLRTQVVN SGS--DMNCN TPTQCYYNTT
                AIPV        GANIAWPLKV NLTRNGHNKV DVVLQNNELM PHGVKTKACV AGVD-QAHCS VESKCYYTNI
                SARS CoV    SPNLAWPLIV TALRA-N--S AVKLQNNELS PVALRQMSCA AGTTQTACTD DNALAYYNNS

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4505       4515       4525       4535       4545       4555
                EMCR        GGKTFMYAYI SNKADLKFVK WEY-EGG-CN TIELDSPCRF MVETPNGPQV KYLYFVKNLN
                229E        GGRAFMYAYV TTKPGMKYVK WEH-DSG-VV TVELEPPCRF VIDTPTGPQI KYLYFVKNLN
                PEDV        GGRTFMYAFI SDKPDLRVVK WEF-DGG-CN TIELEPPRKF LVDSPNGAQI KYLYFVRNLN
                TGEV        SGKSFMYAFI ASDNNLKYVK WES-NND-II PIELEAPLRF YVDGANGPEV KYLYFVKNLN
                OC43        NNGKIVYAIL SDVDGLKYTK ILKDDGN-FV VLELDPPCKF TVQDAKGLKI KYLYFVKGCN
                BoCoV       YNGKIVYAIL SDVDGLKYTK ILKDDGN-FV VLELDPPCKF TVQDVKGLKI KYLYFVKGCN
                MHV         GMGKIVYAIL SDCDGLKYTK IVKEDGN-CV VLELDPPCKF SVQDVKGLKI KYLYFVKGCN
                AIPV        SGNSVVAAIT SSNPNLKVAS FLNEAGN-QI YVDLDPPCKF GMKVGVKVEV VYLYFIKNTR
                SARS CoV    KGGRFVLALL SDHQDLKWAR FPKSDGTGTI YTELEPPCRF VTDTPKGPKV KYLYFIKGLN

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4565       4575       4585       4595       4605       4615
                EMCR        TLRRGAVLGF IGATIRLQAG -KQTELAVNS GLLTACAFSV DPATTYLEAV KHGAKPVSNC
                229E        NLRRGAVLGY IGATVRLQAG -KQTEFVSNS HLLTHCSFAV DPAAAYLDAV KQGAKPVGNC
                PEDV        TLRRGAVLGY IGATVRLQAG -KQTEQAINS SLLTLCAFAV DPAKTYIDAV KSGHKPVGNC
                TGEV        TLRRGAVLGY IGATVRLQAG -KPTEHPSNS SLLTLCAFSP DPAKAYVDAV KRGMQPVNNC
                OC43        TLARGWVVGT ISSTVRLQAG -TATEYASNS SILSLCAFSV DPKKTYLDFI QQGGTPIANC
                BoCoV       TLARGWVVGT ISSTVRLQAG -TATEYASNS SILSLCAFSV DPKKTYLDFI QQGGTPIANC
                MHV         TLARGWVVGT LSSTVRLQAG -TATEYASNS AIRSLCAFSV DPKKTYLDYI QQGGAPVTNC
                AIPV        SIVRGMVLGA ISNVVVLQSK GHETEEVDAV GILSLCSFAV DPADTYCKYV AAGNQPLGNC
                SARS CoV    NLNRGMVLGS LAATVRLQAG -NATEVPANS TVLSFCAFAV DPAKAYKDYL ASGGQPITNC
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       4625       4635       4645       4655       4665       4675
EMCR                IKMLSNGAGN GQAITTSVDA NTNQDSYGGA SICLYCRAHV PHP----SMD GYCKFKGKCV
229E                VKMLTNGSGS GQAITCTIDS NTTQDTYGGA SVCIYCRAHV AHP----TMD GFCQYKGKWV
PEDV                VKMLANGSGN GQAVTNGVEA STNQDSYGGA SVCLYCRAHV EHP----SMD GFCRLKGKYV
TGEV                VKMLSNGAGN GMAVTNGVEA NTQQDSYGGA SVCIYCRCHV EHP----AID GLCRYKGKFV
OC43                VKMLCDHAGT GMAITVKPDA TTSQDSYGGA SVCIYCRARV EHP----DVD GLCKLRGKFV
BoCoV               VKMLCDHAGT GMAITVKPDA TTSQDSYGGA SVCIYCRARV EHP----DVD GLCKLRGKFV
MHV                 VKMLCDHAGT GMAITIKPEA TTNQDSYGGA SVCIYCRSRV EHP----DVD GLCKLRGKFV
AIPV                VKMLTVHNGS GFAITSKPSP TPDQDSYGGA SVCLYCRAHI AHPGSVGNLD GRCQFKGSFV
SARS CoV            VKMLCTHTGT GQAITVTPEA NMDQESFGGA SCCLYCRCHI DHP----NPK GFCDLKGKYV

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       4685       4695       4705       4715       4725       4735
EMCR                QVPIGCL-DP IRFCLENNVC NVCGCWLGHG CACDRTTIQS ---------- -VDISYLNEQ
229E                QVPIGTN-DP IRFCLENTVC KVCGCWLNHG CTCDRTAIQS ---------- -FDNSYLNES
PEDV                QVPLGTV-DP IRFVLENDVC KVCGCWLSNG CTCDRSIMQS ---------- -T--------
TGEV                QIPTGTQ-DP IRFCIENEVC VVCGCWLNNG CMCDRTSMQS F--------- TVDQSYLNEC
OC43                QVPVGIK-DP VSYVLTHDVC RVCGFWRDGS CSCVSTDTTV Q--------- SKDT------
BoCoV               QVPVGIK-DP VSYVLTHDVC QVCGFWRDGS CSCVSTDTTV Q--------- SKDTNFLNGF
MHV                 QVPLGIK-DP VSYVLTHDVC QVCGFWRDGS CSCVGTGSQF Q--------- SKDTNFLNGF
AIPV                QIPTTEK-DP VGFCLRNKVC TVCQCWIGYG CQCDSLRQPK SSVQSVAGAS DFDKNYLNGY
SARS CoV            QIPTTCANDP VGFTLRNTVC TVCGMWKGYG CSCDQLREPL M--------- QSADASTFLN

                    ....|..
                       4745
EMCR                GVLVQLD
229E                GALVPLD
PEDV                -------
TGEV                GVLVQLD
OC43                -------
BoCoV               GVRV---
MHV                 GVQV---
AIPV                GVAVRLG
SARS CoV            GFAV---
```

c. Putative orf 1b

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                          5         15         25         35         45         55
EMCR                ---------- ---------- RARGSSAARL EPCN-GTDID KCVRAFDIYN KNVSFLGKCL
229E                ---------- ---------- ---------- EPCN-GTDID YCVRAFDVYN KDASFIGKNL
PEDV                ---------- -----YGLFK RVRGSSAARL EPCN-GTDTQ HVYRAFDIYN KDVACLGKFL
TGEV                ---------- ---------- ---------- EPCN-GTDPD HVSRAFDIYN KDVACIGKFL
BoCoV               ---------- ------FFKR VRGTSVDARL VPCASGLSTD VQLRAFDICN ASVAGIGLHL
OC43                ---------- ------FFKR VRGTSVDARL VPCASGLSTD VQLRAFDICN ASVAGIGLHL
MHV                 LFLCRHRLPV SVKRHELFKR VRGTSVNARL VPCASGLDTD VQLRAFDICN ANRAGIGLYY
AIPV                ---------- ---------- ---------- ---------- ---------- -----MFQNL
SARS CoV            ---------- ---------- ---------- TPCGTGTSTD VVYRAFDIYN EKVAGFAKFL

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                          65         75         85         95        105        115
EMCR                KMNCVRFKNA --------DL KDGYFVIKRC TKSVMEHEQS MYNLLNFSGA LAEHDFFTWK
229E                KSNCVRFKNV --------DK DDAFYIVKRC IKSVMDHEQS MYNLLKGCNA VAKHDFFTWH
PEDV                KVNCVRLKNL --------DK HDAFYVVKRC TKSAMEHEQS IYSRLEKCGA IAEHDFFTWK
TGEV                KTNCSRFRNL --------DK HDAYYIVKRC TKTVMDHEQV CYNDLKDSGA VAEHDFFTYK
BoCoV               KVNCCRFQRV --DENG--DK LDQFFVVKRT DLTIYNREME CYERVKDCKF VAEHDFFTFD
OC43                KVNCCRFQRV --DENG--DK LDQFFVVKRT DLTIYNREME CYERVKDCKF VAEHDFFTFD
MHV                 KVNCCRFQRA --DEDG--NT LDKFFVIKRT NLEVYNKEKE CYELTKECGV VAEHEFFTFD
AIPV                KRNCARFQEL RDTEDGNLEY LDSYFVVKQT TPSNYEHEKS CYEDLKS-EV TADHDFFVFN
SARS CoV            KTNCCRFQEK --DEEG--NL LDSYFVVKRH TMSNYQHEET IYNLVKDCPA VAVHDFFKFR

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         125        135        145        155        165        175
EMCR                DGRVIYGNVS RHNLTKYTMM DLVYAMRNFD EQNCDVLKEV LVLTGCCDNS ----YFDSKG
229E                EGRTIYGNVS RQDLTKYTMM DLCFALRNFD EKDCEVFKEI LVLTGCCSTD ----YFEMKN
PEDV                DGRAIYGNVC RKDLTEYTMM DLCYALRNFD ENNCDVLKSI LIKVGACEES ----YFNNKV
TGEV                EGRCEFGNVA RRNLTKYTMM DLCYAIRNFD EKNCEVLKEI LVTVGACTEE ----FFENKD
BoCoV               VEGSRVPHIV RKDLTKYTML DLCYALRHFD RNDCMLLCDI LSIYAGCEQS ----YFTKKD
OC43                VEGSRVPHIV RKDLTKYTML DLCYALRHFD RNDCMLLCDI LSIYAGCEQS ----YFTKKD
MHV                 VEGSRVPHIV RKDLSKYTML DLCYALRHFD RNDCSTLKEI LLTYAECDES ----YFQKKD
AIPV                KN---IYNIS RQRLTKYTMM DFCYALRHFD PKDCEVLKEI LVTYGCIEDY HPKWFEENKD
SARS CoV            VDGDMVPHIS RQRLTKYTMA DLVYALRHFD EGNCDTLKEI LVTYNCCDDD ----YFNKKD
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      185        195        205        215        225        235
EMCR                WYDPVENEDI HRVYASLGKI VARAMLKCVA LCDAMVAKGV VGVLTLDNQD LNGNFYDFGD
229E                WFDPIENEDI HRVYAALGKV VANAMLKCVA FCDEMVLKGV VGVLTLDNQD LNGNFYDFGD
PEDV                WFDPVENEDI HRVYALLGTI VARAMLKCVK FCDAMVEQGI VGVVTLDNQD LNGDFYDFGD
TGEV                WFDPVENEAI HEVYAKLGPI VANAMLKCVA FCDAIVEKGY IGVITLDNQD LNGNFYDFGD
BoCoV               WYDFVENPDI INVYKKLGPI FNRALVSATE FADKLVEVGL VGILTLDNQD LNGKWYDFGD
OC43                WYDFVENPDI INVYKKLGPI FNRALVSATE FADKLVEVGL VGVLTLDNQD LNGKWYDFGD
MHV                 WYDFVENSDI INVYKKLGPI FNRALLNTAK FADTLVEAGL VGVLTLDNQD LYGQWYDFGD
AIPV                WYDPIENSKY YVMLAKMGPI VRRALLNAIE FGNLMVEKGY VGVITLDNQD LNGKFYDFGD
SARS CoV            WYDFVENPDI LRVYANLGER VRQSLLKTVQ FCDAMRDAGI VGVLTLDNQD LNGNWYDFGD

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      245        255        265        275        285        295
EMCR                FVVSLPNMGV PCCTSYYSYM MPIMGLTNCL ASECFVKSDI FGSDFKTFDL LKYDFTEHKE
229E                FVLCPPGMGI PYCTSYYSYM MPVMGMTNCL ASECFMKSDI FGQDFKTFDL LKYDFTEHKE
PEDV                FTCSIKGMGV PICTSYYSYM MPVMGMTNCL ASECFVKSDI FGEDFKSYDL LEYDFTEHKT
TGEV                FVKTAPGFGC ACVTSYYSYM MPLMGMTSCL ESENFVKSDI YGSDYKQYDL LAYDFTEHKE
BoCoV               YVIAAPGCGV AIADSYYSYM MPMLTMCHAL DCELYVNNAY R-----LFDL VQYDFTDYKL
OC43                YVIAAPGCGV AIADSYYSYI MPMLTMCHAL DCELYVNNAY R-----LFDL VQYDFTDYKL
MHV                 FVKTVPGCGV AVADSYYSYM MPMLTMCHAL DSELFINGTY R-----EFDL VQYDFTDFKL
AIPV                FQKTAPGAGV PVFDTYYSYM MPIIAMTDAL APERYFEYDV HKG-YKSYDL LKYDYTEEKQ
SARS CoV            FVQVAPGCGV PIVDSYYSLL MPILTLTRAL AAESHMDADL AKP-LIKWDL LKYDFTEERL

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      305        315        325        335        345        355
EMCR                NLFNKYFKHW SFDYHPNCSD CYDDMCVIHC ANFNTLFATT IPGTAFGPLC RKVFIDGVPL
229E                VLFNKYFKYW GQDYHPDCVD CHDEMCILHC SNFNTLFATT IPNTAFGPLC RKVFIDGVPV
PEDV                ALFNKYFKYW GLQYHPNCVD CSDEQCIVHC ANFNTLFSTT IPITAFGPLC RKCWIDGVPL
TGEV                YLFQKYFKYW DRTYHPNCSD CTSDECIIHC ANFNTLFSMT IPMTAFGPLV RKVHIDGVPV
BoCoV               ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV RQIFVDGVPF
OC43                ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV RQIFVDGVPF
MHV                 ELFNKYFKYW SMTYHPNTCE CEDDRCIIHC ANFNILFSMV LPKTCFGPLV RQIFVDGVPF
AIPV                ELFQKYFKYW DQEYHPNCRD CSDDRCLIHC ANFNILFSTL IPQTSFGNLC RKVFVDGVPF
SARS CoV            CLFDRYFKYW DQTYHPNCIN CLDDRCILHC ANFNVLFSTV FPPTSFGPLV RKIFVDGVPF

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      365        375        385        395        405        415
EMCR                VTTAGYHFKQ LGLVWNKDVN THSVRLTITE LLQFVTDPSL IIASSPALVD QRTICFSVAA
229E                VATAGYHFKQ LGLVWNKDVN THSTRLTITE LLQFVTDPTL IVASSPALVD KRTVCFSVAA
PEDV                VTTAGYHFKQ LGIVWNNDLN LHSSRLSINE LLQFCSDPAL LIASSPALVD QRTVCFSVAA
TGEV                VVTAGYHFKQ LGIVWNLDVK LDTMKLSMTD LLRFVTDPTL LVASSPALLD QRTVCFSIAA
BoCoV               VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD LRTCCFSVAA
OC43                VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD LRTCCFSVAA
MHV                 VVSIGYHYKE LGVVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALLD LRTCCFSVAA
AIPV                IATCGYHSKE LGVIMNQDNT MSFSKMGLSQ LMQFVGDPAL LVGTSNNLVD LRTSCFSVCA
SARS CoV            VVSTGYHFRE LGVVHNQDVN LHSSRLSFKE LLVYAADPAM HAASGNLLLD KRTTCFSVAA

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      425        435        445        455        465        475
EMCR                LSTGLTNQVV KPGHFNEEFY NFLRLRGFFD EGSELTLKHF FFAQNGDAAV KDFDFYRYNK
229E                LSTGLTSQTV KPGHFNKEFY DFLRSQGFFD EGSELTLKHF FFTQKGDAAI KDFDYYRYNR
PEDV                LGTGMTNQTV KPGHFNKEFY DFLLEQGFFS EGSELTLKHF FFAQKVDAAV KDFDYYRYNR
TGEV                LSTGITYQTV KPGHFNKDFY DFITERGFFE EGSELTLKHF FFAQGGEAAM TDFNYYRYNR
BoCoV               ITSGVKFQTV KPGNFNQDFY DFILSKGLLK EGSSVDLKHF FFTQDGNAAI TDYNYYKYNL
OC43                ITSGVKFQTV KPGNFNQDFY DFVLSKGLLK EGSSVDLKHF FFTQDGNAAI TDYNYYKYNL
MHV                 ITSGVKFQTV KPGNFNQDFY EFILSKGLLK EGSSVDLKHF FFTQDGNAAI TDYNYYKYNL
AIPV                LTSGITHQTV KPGHFNKDFY DFAEKAGMFK EGSSIPLKHF FYPQTGNAAI NDYDYYRYNR
SARS CoV            LTNNVAFQTV KPGNFNKDFY DFAVSKGFFK EGSSVELKHF FFAQDGNAAI SDYDYYRYNL

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      485        495        505        515        525        535
EMCR                PTILDICQAR VTYKIVSRYF DIYEGGCIKA CEVVVTNLNK SAGWPLNKFG KASLYYESIS
229E                PTMLDIGQAR VAYQVAARYF DCYEGGCITS REVVVTNLNK SAGWPLNKFG KAGLYYESIS
PEDV                PTVLDICQAR VVYQIVQRYF DIYEGGCITA KEVVVTNLNK SAGYPLNKFG KAGLYYESLS
TGEV                VTVLDICQAQ FVYKIVGKYF ECYDGGCINA REVVVTNYDK SAGYPLNKFG KARLYYETLS
BoCoV               PTMVDIKQLL FVLEVVYKYF EIYDGGCIPA AQVIVNNYDK SAGYPFNKFG KARLYYEALS
OC43                PTMVDIKQLL FVLEVVYKYF EIYDGGCIPA SQVIVNNYDK SAGYPFNKFG KARLYYEALS
MHV                 PTMVDIKQLL FVLEVVNKYF EIYDGGCIPA TQVIVNNYDK SAGYPFNKFG KARLYYEALS
AIPV                PTMFDICQLL FCLEVTSKYF ECYEGGCIPA SQVVVNNLDK SAGYPFNKFG KARLYYE-MS
SARS CoV            PTMCDIRQLL FVVEVVDKYF DCYDGGCINA NQVIVNNLDK SAGFPFNKWG KARLYYDSMS

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                      545        555        565        575        585        595
EMCR                YEEQDALFAL TKRNVLPTMT QLNLKYAISG KERARTVGGV SLLSTMTTRQ YHQKHLKSIV
229E                YEEQDAIFSL TKRNILPTMT QLNLKYAISG KERARTVGGV SLLATMTTRQ FHQKCLKSIV
PEDV                YEEQDELYAY TKRNILPTMT QLNLKYAISG KERARTVGGV SLLSTMTTRQ YHQKHLKSIV
TGEV                YEEQDALFAL TKRNVLPTMT QMNLKYAISG KARARTVGGV SLLSTMTTRQ YHQKHLKSIA
BoCoV               FEEQDEIYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM FHQKCLKSIA
```

```
OC43        FEEQDEIYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM FHQKCLKSIA
MHV         FEEQDEVYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM FHQKCLKSIA
AIPV        LEEQDQLFEI TKKNVLPTIT QMNLKYAISA KNRARTVAGV SILSTMTNRQ FHQKILKSIV
SARS CoV    YEDQDALFAY TKRNVIPTIT QMNLKYAISA KNRARTVAGV SICSTMTNRQ FHQKLLKSIA

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                605        615        625        635        645        655
EMCR        NTRNATVVIG TTKFYGGWNN MLRTLIDGVE NPMLMGWDYP KCDRALPNMI RMISAMVLGS
229E        ATRNATVVIG TTKFYGGWDN MLKNLMADVD DPKLMGWDYP KCDRAMPSMI RMLSAMILGS
PEDV        NTRGASVVIG TTKFYGGWDN MLKNLIDGVE NPCLMGWDYP KCDRALPNMI RMISAMILGS
TGEV        ATRNATVVIG STKFYGGWDN MLKNLMRDVD NGCLMGWDYP KCDRALPNMI RMISAMILGS
BoCoV       ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD NPVLMGWDYP KCDRAMPNIL RMASAMILGS
OC43        ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD NPVLMGWDYP KCDRAMPNLL RIVSSLVLAR
MHV         ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD SPVLMGWDYP KCDRAMPNIL RIISSLVLAR
AIPV        NTRNASVVIG TTKFYGGWDN MLRNLIQGVE DPILMGWDYP KCDRAMPNLL RIAASLVLAR
SARS CoV    ATRGATVVIG TSKFYGGWHN MLKTVYSDVE TPHLMGWDYP KCDRAMPNML RIMASLVLAR

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                665        675        685        695        705        715
EMCR        KHVNCCTVTD RFYRLGNELA QVLTEVVYSN GGFYFKPGGT TSGDASTAYA NSIFNIFQAV
229E        KHVTCCTASD KFYRLSNELA QVLTEVVYSN GGFYFKPGGT TSGDATTAYA NSVFNIFQAV
PEDV        KHTTCCSSTD RFFRLCNELA QVLTEVVYSN GGFYLKPGGT TSGDATTAYA NSVFNIFQAV
TGEV        KHVGCCTHND RFYRLSNELA QVLTEVVHCT GGFYFKPGGT TSGDGTTAYA NSAFNIFQAV
BoCoV       KHEACCSQSD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA NSVFNICQAV
OC43        KHETCCSQSD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA NSVFNICQAV
MHV         KHDSCCSHTD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA NSVFNICQAV
AIPV        KHTNCCSWSE RIYRLYNECA QVLSETVLAT GGIYVKPGGT SSGDATTAYA NSVFNIIQAT
SARS CoV    KHNTCCNLSH RFYRLANECA QVLSEMVMCG GSLYVKPGGT SSGDATTAYA NSVFNICQAV

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                725        735        745        755        765        775
EMCR        SSNINRLLSV PSDSCNNVNV RDLQRRLYDN CYRLTSVEES FIDDYYGYLR KHFSMMILSD
229E        SSNINCVLSV NSSNCNNFNV KKLQRQLYDN CYRNSNVDES FVDDFYGYLQ KHFSMMILSD
PEDV        SANVNKLLSV DSNVCHNLEV KQLQRKLYEC CYRSTIVDDQ FVVEYYGYLR KHFSMMILSD
TGEV        SANVNKLLGV DSNACNNVTV KSIQRKIYDN CYRSSSIDEE FVVEYFSYLR KHFSMMILSD
BoCoV       SANVCALMSC NGNKIEDLSI RALQKRLYSH VYRSDMVDST FVTEYYEFLN KHFSMMILSD
OC43        SANVCALMSC NGNKIEDLSI RALQKRLYSH VYRSDKVDST FVTEYYEFLN KHFSMMILSD
MHV         SANVCSLMAC NGHKIEDLSI RELQKRLYSN VYRADHVDPA FVSEYYEFLN KHFSMMILSD
AIPV        SANVARLLSV ITRDIVYDNI KSLQYELYQQ VYRRVNFDPA FVEKFYSYLC KNFSLMILSD
SARS CoV    TANVNALLST DGNKIADKYV RNLQHRLYEC LYRNRDVDHE FVDEFYAYLR KHFSMMILSD

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                785        795        805        815        825        835
EMCR        DGVVCYNKDY AELGYIADIS AFKATLYYQN NVFMSTSKCW VEEDLTKGPH EFCSQHTMQI
229E        DSVVCYNKTY AGLGYIADIS AFKATLYYQN GVFMSTAKCW TEEDLSIGPH EFCSQHTMQI
PEDV        DGVVCYNNDY ASLGYVADLN AFKAVLYYQN NVFMSASKCW IEPDINKGPH EFCSQHTMQI
TGEV        DGVVCYNKDY ADLGYVADIN AFKATLYYQN NVFMSTSKCW VEPDLSVGPH EFCSQHTLQI
BoCoV       DGVVCYNSDY ASKGYIANIS AFQQVLYYQN NVFMSESKCW VENDINNGPH EFCSQHTMLV
OC43        DGVVCYNSDY ASKGYIANIS AFQQVLYYQN NVFMSESKCW VEHDINNGPH EFCSQHTMLV
MHV         DGVVCYNSEF ASKGYIANIS AFQQVLYYQN NVFMSEAKCW VETDIEKGPH EFCSQHTMLV
AIPV        DGVVCYNNTL AKQGLVADIS GFREVLYYQN NVFMADSKCW VEPDLEKGPH EFCSQHTMLV
SARS CoV    DAVVCYNSNY AAQGLVASIK NFKAVLYYQN NVFMSEAKCW TETDLTKGPH EFCSQHTMLV

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                845        855        865        875        885        895
EMCR        VDKDGTYYLP YPDPSRILSA GVFVDDVVKT DAVVLLXRYV SLAIDAYPLS KHPNSEYRKV
229E        VDENGKYYLP YPDPSRIISA GVFVDDITKT DAVILLERYV SLAIDAYPLS KHPKPEYRKV
PEDV        VDKEGTYYLP YPDPSRILSA GVFVDDVVKT DAVVLLERYV SLAIDAYPLS KHENPEYKKV
TGEV        VGPDGDYYLP YPDPSRILSA GVFVDDIVKT DNVIMLERYV SLAIDAYPLT KHPKPAYQKV
BoCoV       KMDGDDVYLP YPVPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV YHENEEYQKV
OC43        KMDGDDVYLP YPNPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV YHENEEYQKV
MHV         KMDGDEVYLP YPDPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV YHENEEYQKV
AIPV        EVDGEPKYLP YPDPSRILGA CVFVDDVDKT EPVAVMERYI ALAIDAYPLV HHENPEYQNV
SARS CoV    KQGDDYVYLP YPDPSRILGA GCFVDDIVKT DGTLMIERFV SLAIDAYPLT KHPNQEYADV

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                905        915        925        935        945        955
EMCR        FYVLLDWVKH LNKNLNEGVL ESFSVTLLDN QEDKFWCEDF YASMYENSTI LQAAGLCVVC
229E        FYALLDWVKH LNKTLNEGVL ESFSVTLLDE HESKFWDESF YASMYEKSTV LQAAGLCVVC
PEDV        FYVLLDWVKH LYKTLNAGVL ESFSVTLLED STAKFWDESF YANMYEKSAV LQSAGLCVVC
TGEV        FYTLLDWVKH LQKNLNAGVL DSFSVTMLEE GQDKFWSEEF YASLYEKSTV LQAAGMCVVC
BoCoV       FRVYLEYIKK LYNELGNQIL DSYSVILSTC DGQKFTDESF YKNMYLRSAV MQSVGACVVC
OC43        FRVYLAYIKK LYNDLGNQIL DSYSVILSTC DGQKFTDESF YKNMYLRSAV MQSVGACVVC
MHV         FRVYLEYIKK LYNDLGNQIL DSYSVILSTC DGQKFTDETF YKNMYLRSAV MQSVGACVVC
AIPV        FFVLLAYIRK LYQELSQNML MDYSFVMDID KGSKFWEQEF YENMYRAPTT LQSCGVCVVC
SARS CoV    FHLYLQYIRK LHDELTGHML DMYSVMLTND NTSRYWEPEF YEAMYTPHTV LQAVGACVLC

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                965        975        985        995        1005       1015
EMCR        GSQTVLRCGD CLRKPMLCTK CAYDHVFGTD HKFILAITPY VCNASGCGVS DVKKLYLGGL
```

```
                229E        GSQTVLRCGD CLRRPMLCTK CAYDHVFGTD HKFILAITPY VCNTSGCNVN DVTKLYLGGL
                PEDV        GSQTVLRCGD CLRRPMLCTK CAYDHVIGTT HKFILAITPY VCCASDCGVN DVTKLYLGGL
                TGEV        GSQTVLRCGD CLRRPLLCTK CAYDHVMGTK HKFIMSITPY VCSFNGCNVN DVTKLFLGGL
                BoCoV       SSQTSLRCGS CIRKPLLCCK CCYDHVMATD HKYVLSVSPY VCNAPGCDVN DVTKLYLGGM
                OC43        SSQTSLRCGS CIRKPLLCCK CCYDHVMATD HKYVLSVSPY VCNAPGCDVN DVTKLYLGGM
                MHV         SSQTSLRCGS CIRKPLLCCK CAYDHVMSTD HKYVLSVSPY VCNSPGCDVN DVTKLYLGGM
                AIPV        NSQTILRCGN CIRKPFLCCK CCYDHVMHTD HKNVLSINPY ICSQLGCGEA DVTKLYLGGM
                SARS CoV    NSQTSLRCGA CIRRPFLCCK CCYDHVISTS HKLVLSVNPY VCNAPGCDVT DVTQLYLGGM

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1025       1035       1045       1055       1065       1075
                EMCR        NYYCTNHKPQ LSFPLCSAGN IFGLYKNSAT GSLDVEVFNR LATSDWTDVR DYKLANDVKD
                229E        NYYCVDHKPH LSFPLCSAGN VFGLYKSSAL GSMDIDVFNK LSTSDWSDIR DYKLANDAKE
                PEDV        SYWCHEHKPR LAFPLCSAGN VFGLYKNSAT GSPDVEDFNR IATSDWTDVS DYRLANDVKD
                TGEV        SYYCMNHKPQ LSFPLCANGN VFGLYKSSAV GSEAVEDFNK LAVSDWTNVE DYKLANNVKE
                BoCoV       SYYCEDHKPQ YSFKLVMNGM VFGLYKQSCT GSPYIDDFNR IASCKWTDVD DYILANECTE
                OC43        SYYCEDHKPQ YSFKLVMNGL VFGLYKQSCT GSPYIDDFNR IASCKWTDVD DYILANECTE
                MHV         SYYCEDHKPQ YSFKLVMNGM VFGLYKQSCT GSPYIEDFNK IASCKWTEVD DYVLANECTE
                AIPV        SYFCGNHKPK LSIPLVSNGT VFGIYRANCA GSENVDDFNQ LATTNWSIVE PYILANRCSD
                SARS CoV    SYYCKSHKPP ISFPLCANGQ VFGLYKNTCV GSDNVTDFNA IATCDWTNAG DYILANTCTE

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1085       1095       1105       1115       1125       1135
                EMCR        TLRLFAAETI KAKEESVKSS YAFATLKEVV GPKELLLSWE SGKVKPPLNR NSVFTCFQIS
                229E        SLRLFAAETV KAKEESVKSS YAYATLKEIV GPKELLLLWE SGKAKPPLNR NSVFTCFQIT
                PEDV        SLRLFAAETI KAKEESVKSS YACATLHEVV GPKELLLKWE VGRPKPPLNR NSVFTCYHIT
                TGEV        SLKIFAAETV KAKEESVKSE YAYAVLKEVI GPKEIVLQWE ASKTKPPLNR NSVFTCFQIS
                BoCoV       RLKLFAAETQ KATEEAFKQS YASATIQEIV SERELILSWE IGKVKPPLNK NYVFTGYHFT
                OC43        RLKLFAAETQ KATEEAFKQS YASATIQEIV SERELILSWE IGKVKPPLNK NYVFTGYHFT
                MHV         RLKLFAAETQ KATEESFKQC YASATIREIV SDRELILSWE IGKVRPPLNK NYVFTGYHFT
                AIPV        SLRRFAAETV KATEELHKQQ FASAEVREVF SDRELILSWE PGKTRPPLNR NYVFTGYHFT
                SARS CoV    RLKLFAAETL KATEETFKLS YGIATVREVL SDRELHLSWE VGKPRPPLNR NYVFTGYRVT

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1145       1155       1165       1175       1185       1195
                EMCR        KDSKFQIGEF IFEKVEYGSD TVTYKSTVTT KLVPGMIFVL TSHNVQPLRA PTIANQEKYS
                229E        KDSKFQVGEF VFEKVDYGSD TVTYKSTATT KLVPGMLFIL TSHNVAPLRA PTMANQEKYS
                PEDV        KNTKFQIGEF VFEKAEYDND AVTYKTTATT KLVPGMVFVL TSHNVQPLRA PTIANQERYS
                TGEV        KDTKIQLGEF VFEQSEYGSD SVYYKSTSTY KLTPGMIFVL TSHNVSPLKA PILVNQEKYN
                BoCoV       KNGKTVLGEY VFDKSEL-TN GVYYRATTTY KLSVGDVFVL TSHSVANLSA PTLVPQENYS
                OC43        KNGKTVLGEY VFDKSEL-TN GVYYRATTTY KLSVGDVFVL TSHSVANLSA PTLVPQENYS
                MHV         SNGKTVLGEY VFDKSEL-TN GVYYRATTTY KLSVGDVFIL TSHAVSSLSA PTLVPQENYT
                AIPV        RTSKVQLGDF TFEKGEG-KD VVYYKATSTA KLSVGDIFVL TSHNVVSLVA PTLCPQQTFS
                SARS CoV    KNSKVQIGEY TFEKGDY-GD AVVYRGTTTY KLNVGDYFVL TSHTVMPLSA PTLVPQEHYV

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1205       1215       1225       1235       1245       1255
                EMCR        SIYKLHPAFN VSDAYANLVP YYQLIGKQKI TTIQGPPGSG KSHCSIGLGL YYPGARIVFV
                229E        TIYKLHPSFN VSDAYANLVP YYQLIGKQRI TTIQGPPGSG KSHCSIGIGV YYPGARIVFT
                PEDV        TIHKLHPAFN IPEAYSSLVP YYQLIGKQKI TTIQGPPGSG KSHCVIGLGL YYPGARIVFT
                TGEV        TISKLYPVFN IAEAYNTLVP YYQMIGKQKF TTIQGPPGSG KSHCVIGLGL YYPQARIVYT
                BoCoV       SIR-FASVYS VLETFQNNVV NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV YYCTARVVYT
                OC43        SIR-FASVYS VLETFQNNVV NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV YYCTARVVYT
                MHV         SIR-FASVYS VPETFQNNVP NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV YYCTARVVYT
                AIPV        RFVNLRPNVM VPECFVNNIP LYHLVGKQKR TTVQGPPGSG KSHFAIGLAV YFSSARVVFT
                SARS CoV    RITGLYPTLN ISDEFSSNVA NYQKVGMQKY STLQGPPGTG KSHFAIGLAL YYPSARIVYT

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1265       1275       1285       1295       1305       1315
                EMCR        ACAHAAVDSL CAKAMTVYSI DKCTRIIPAR ARVECYSGFK PNNTSAQYIF STVNALPECN
                229E        ACSHAAVDSL CAKAVTAYSV DKCTRIIPAR ARVECYSGFK PNNNSAQYVF STVNALPEVN
                PEDV        ACSHAAVDSL CVKASTAYSN DKCSRIIPQR ARVECYDGFK SNNTSAQYLF STVNALPECN
                TGEV        ACSHAAVDAL CEKAAKNFNV DRCSRIIPQR IRVDCYTGFK PNNTNAQYLF CTVNALPEAS
                BoCoV       AASHAAVDAL CEKAYKFLNI NDCTRIVPAK VRVECYDKFK INDTTRKYVF TTINALPEMV
                OC43        AASHAAVDAL CEKAYKFLNI NDCTRIVPAK VRVECYDKFK INDTTRKYVF TTINALPEMV
                MHV         AASHAAVDAL CEKAYKFLNI NDCTRIVPAR VRVDCYDKFK VNDTTRKYVF TTINALPELV
                AIPV        ACSHAAVDAL CEKAFKFLKV DDCTRIVPQR TTVDCFSKFK ANDTGKKYIF STINALPEVS
                SARS CoV    ACSHAAVDAL CEKALKYLPI DKCSRIIPAR ARVECFDKFK VNSTLEQYVF CTVNALPETT

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1325       1335       1345       1355       1365       1375
                EMCR        ADIVVVDEVS MCTNYDLSVI NQRLSYKHIV YVGDPQQLPA PRVMITKGVM EPVDYNVVTQ
                229E        ADIVVVDEVS MCTNYDLSVI NQRISYKHIV YVGDPQQLPA PRVLISKGVM EPIDYNVVTQ
                PEDV        ADIVVVDEVS MCTNYDLSVI NQRISYRHVV YVGDPQQLPA PRVMISRGTL EPKDYNVVTQ
                TGEV        CDIVVVDEVS MCTNYDLSVI NSRLSYKHIV YVGDPQQLPA PRTLINKGVL QPQDYNVVTK
                BoCoV       TDIVVVDEVS MLTNYELSVI NARIRAKHYV YIGDPAQLPA PRVLLSKGTL EPKYFNTVTK
                OC43        TDIVVVDEVS MLTNYELSVI NARIRAKHYV YIGDPAQLPA PRVLLSKGTL EPKYFNTVTK
                MHV         TDIIVVDEVS MLTNYELSVI NSRVRAKHYV YIGDPAQLPA PRVLLNKGTL EPRYFNSVTK
                AIPV        CDILLVDEVS MLTNYELSFI NGKINYQYVV YVGDPAQLPA PRTLLN-GSL SPKDYNVVTN
                SARS CoV    ADIVVFDEIS MATNYDLSVV NARLRAKHYV YIGDPAQLPA PRTLLTKGTL EPEYFNSVCR
```

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1385       1395       1405       1415       1425       1435
EMCR                RMCAIGPDVF LHKCYRCPAE IVNTVSELVY ENKFVPVKPA SKQCFKIFFK G---NVQVDN
229E                RMCAIGPDVF LHKCYRCPAE IVNTVSELVY ENKFVPVKEA SKQCFKIFER G---SVQVDN
PEDV                RMCALKPDVF LHKCYRCPAE IVRTVSEMVY ENQFIPVHPD SKQCFKIFCK G---NVQVDN
TGEV                RMCTLGPDVF LHKCYRCPAE IVKTVSALVY ENKFVPVNPE SKQCFKMFVK G---QVQIES
BoCoV               LMCCLGPDIF LGTCYRCPKE IVDTVSALVY ENKLKAKNES SSLCFKVYYK G---VTTHES
OC43                LMCCLGPDIF LGTCYRCPKE IVDTVSALVY ENKLKAKNES SSLCFKVYYK G---VTTHES
MHV                 LMCCLGPDIF LGTCYRCPKE IVDTVSALVY HNKLKAKNDN SSMCFKVYYK G---QTTHES
AIPV                LMVCVKPDIF LAKCYRCPKE IVDTVSTLVY DGKFIANNPE SRECFKVIVN NGNSDVGHES
SARS CoV            LMKTIGPDMF LGTCRRCPAE IVDTVSALVY DNKLKAHKDK SAQCFKMFYK G---VITHDV

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1445       1455       1465       1475       1485       1495
EMCR                GSSINRKQLE IVKLFLVKNP SWSKAVFISP YNSQNYVASR FLGLQIQTVD SSQGSEYDYV
229E                GSSINRRQLD VVKRFIHKNS TWSKAVFISP YNSQNYVAAR LLGLQTQTVD SAQGSEYDYV
PEDV                GSSINRRQLD VVRMFLAKNP RWSKAVFISP YNSQNYVASR LLGLQIQTVD SSQGSEYDYV
TGEV                NSSINNKQLE VVKAFLAHNP KWRKAVFISP YNSQNYVARR LLGLQTQTVD SAQGSEYDYV
BoCoV               SSAVNMQQIY LINKFLKANP LWHKAVFISP YNSQNFAAKR VLGLQTQTVD SAQGSEYDYV
OC43                SSAVNMQQIY LINKFLKANP LWHKAVFISP YNSQNFAAKR VLGLQTQTVD SAQGSEYDYV
MHV                 SSAVNMQQIY LISKFLKANP SWSNAVFISP YNSQNYVAKR VLGLQTQTVD SAQGSEYDYV
AIPV                GSAYNTTQLE FVKDFVCRNK QWREAIFISP YNAMNQRAYR MLGLNVQTVD SSQGSEYDFV
SARS CoV            SSAINRPQIG VVREFLTRNP AWRKAVFISP YNSQNAVASK ILGLPTQTVD SSQGSEYDYV

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1505       1515       1525       1535       1545       1555
EMCR                IYAQTSDTAH ACNVNRFNVA ITRAKKGIFC VMCDKT-LFD SLKFFEIKHA --DLHSS---
229E                IFAQTSDTAH ACNANRFNVA ITRAKKGIFC IMSDRT-LFD ALKFFEITMT --DLQSE---
PEDV                IYAQTSDTAH ASNVNRFNVA ITRAKKGILC IMCDRS-LFD LLKFFELKLS --DLQAN---
TGEV                IYTQTSDTQH ATNVNRFNVA ITRAKVGILC IMCDRT-MYE NLDFYELKDS KIGLQAKP--
BoCoV               IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSNMQ-LFE ALQFTTLTVD KVPQAVETRV
OC43                IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSNMQ-LFE ALQFTTLTLD KVPQAVETKV
MHV                 IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSSMQ-LFE SLNFSTLTLD KIN---NPRL
AIPV                IFCVTADSQH ALNINRFNVA LTRAKRGILV VMRQRDELYS ALKFTELDSE TSLQG-----
SARS CoV            IFTQTTETAH SCNVNRFNVA ITRAKIGILC IMSDRD-LYD KLQFTSLEIP RRN-VATLQA

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1565       1575       1585       1595       1605       1615
EMCR                -QVCGLFKNC TRTPLNLPPT HAHTFLSLSD QFKTTGDLAV QIGSNN--VC TYEHVISFMG
229E                -SSCGLFKDC ARNPIDLPPS HATTYLSLSD RFKTSGDLAV QIGNNN--VC TYEHVISYMG
PEDV                -EGCGLFKDC SRGDDLLPPS HANTFMSLAD NFKTDQYLAV QIGVNG--PI KYEHVISFMG
TGEV                -ETCGLFKDC SKSEQYIPPA YATTYMSLSD NFKTSDGLAV NIG-TK--DV KYANVISYMG
BoCoV               QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV TYSRLISLMG
OC43                QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV TYSRLISLMG
MHV                 QCTTNLFKDC SRSYAGYHPA HAPSFLAVDD KYKVGGDLAV CLNVAD-SAV TYSRLISLMG
AIPV                ---TGLFKIC NKEFSGVHPA YAVTTKALAA TYKVNDELAA LVNVEAGSEI TYKHLISLLG
SARS CoV            ENVTGLFKDC SKIITGLHPT QAPTHLSVDI KFKTEG-LCV DIPGIP-KDM TYRRLISMMG

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1625       1635       1645       1655       1665       1675
EMCR                FRFDISIPGS HSLFCTRDFA IRNVRGWLGM DVESAHVCGD NIGTNVPLQV GFSNGVNFVV
229E                FRFDVSMPGS HSLFCTRDFA MRHVRGWLGM DVEGAHVTGD NVGTNVPLQV GFSNGVDFVA
PEDV                FRFDINIPNH HTLFCTRDFA MRNVRGWLGF DVEGAHVVGS NVGTNVPLQV GFSNGVDFVV
TGEV                FRFEANIPGY HTLFCTRDFA MRNVRAWLGF DVEGAHVCGD NVGTNVPLQL GFSNGVDFVV
BoCoV               FKLDVTLDGY CKLFITKEEA VKRVRAWVGF DAEGAHATRD SIGTNFPLQL GFSTGIDFVV
OC43                FKLDVTLDGY CKLFITKEEA VKRVRAWVGF DAEGAHATRD SIGTNFPLQL GFSTGIDFVV
MHV                 FKLDLTLDGY CKLFITRDEA IRRVRAWVGF DAEGAHATRD SIGTNFPLQL GFSTGIDFVV
AIPV                FKMSVNVEGC HNMFITRDEA IRNVRGWVGF DVEATHACGT NIGTNLPFQV GFSTGADFVV
SARS CoV            FKMNYQVNGY PNMFITREEA IRHVRAWIGF DVEGCHATRD AVGTNLPLQL GFSTGVNLVA

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1685       1695       1705       1715       1725       1735
EMCR                QTEGCVSTNF GDVIKPVCAK SPPGEQFRHL VPFLRKGQPW LIVRRRIVQM ISDYLSNLSD
229E                QPEGCVLTNT GSVVKPVRAR APPGEQFTHI VPLLRKGQPW SVLRKRIVQM IADFLAGSSD
PEDV                RPEGCVVTES GDYIKPVRAR APPGEQFAHL LPLLKRGQPW DVVRKRIVQM CSDYLANLSD
TGEV                QTEGCVITEK GNSIEVVKAR APPGEQFAHL IPLMRKGQPW HIVRRRIVQM VCDYFDGLSD
BoCoV               EATGLFADRD GYSFKKAVAK APPGEQFKHL IPLMTRGQRW DVVRPRIVQM FADHLIDLSD
OC43                EATGLFADRD GYSFKKAVAK APPGEQFKHL IPLMTRGQRW DVVRPRIVQM FADHLIDLSD
MHV                 EATGMFAERD GYVFKKAVAR APPGEQFKHL VPLMSRGQKW DVVRIRIVQM LSDHLVDLAD
AIPV                TPEGLVDTSI GNNFEPVNSK APPGEQFNHL RVLFKSAKPW HVIRPRIVQM LADNLCNVSD
SARS CoV            VPTGYVDTEN NTEFTRVNAK PPPGDQFKHL IPLMYKGLPW NVVRIKIVQM LSDTLKGLSD

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                       1745       1755       1765       1775       1785       1795
EMCR                ILVFVLWAGS LELTTMRYFV KIGPIKYCY- CGNSATCYNS VSNEYCCFKH ALGCDYVYNP
229E                VLVFVLWAGG LELTTMRYFV KIGAVKHCQ- CGTVATCYNS VSNDYCCFKH ALGCDYVYNP
PEDV                ILIFVLWAGG LELTTMRYFV KIGPSKSCD- CGKVATCYNS ALHTYCCFKH ALGCDYLYNP
TGEV                ILIFVLWAGG LELTTMRYFV KIGRPQKCE- CGKSATCYSS SQSVYACFKH ALGCDYLYNP
BoCoV               CVVLVTWAAN FELTCLRYFA KVGREISCNV STKRATAYNS RTGYYGCWRH SVTCDYLYNP
```

```
OC43        CVVLVTWAAN FELTCLRYFA KVGREISCNV CTKRATVYNS RTGYYGCWRH SVTCDYLYNP
MHV         SVVLVTWAAS FELTCLRYFA KVGKEVVCSV CNKRATCFNS RTGYYGCWRH SYSCDYLYNP
AIPV        CVVFVTWCHG LELTTLRYFV KIGKEQVCS- CGSRATTFNS HTQAYACWKH CLGFDFVYNP
SARS CoV    RVVFVLWAHG FELTSMKYFV KIGPERTCCL CDKRATCFST SSDTYACWNH SVGFDYVYNP

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1805       1815       1825       1835       1845       1855
EMCR        YAFDIQQWGY VGSLSQNHHT FCNIHRNEHD ASGDAVMTRC LAVHDCFVKN VDWTVTYPFI
229E        YVIDIQQWGY VGSLSTNHHA ICNVHRNEHV ASGDAIMTRC LAVYDCFVKN VDWSITYPMI
PEDV        YCIDIQQWGY KGSLSLNHHE HCNVHRNEHV ASGDAIMTRC LAIHDCFVKN VDWSITYPFI
TGEV        YCIDIQQWGY TGSLSMNHHE VCNIHRNEHV ASGDAIMTRC LAIHDCFVKR VDWSIVYPFI
BoCoV       LIVDIQQWGY IGSLSSNHDL YCSVHKGAHV ASSDAIMTRC LAVYDCFCNN INWNVEYPII
OC43        LIVDIQQWGY IGSLSSNHDL YCSVHKGAHV ASSDAIMTRC LAVYDCFCNN INWNVEYPII
MHV         LIVDIQQWGY TGSLTSNHDL ICSVHKGAHV ASSDAIMTRC LAVHDCFCKS VNWSLEYPII
AIPV        LLVDIQQWGY SGNLQFNHDL HCNVHGHAHV ASVDAIMTRC LAINNAFCQD VNWDLTYPHI
SARS CoV    FMIDVQQWGF TGNLQSNHDQ HCQVHGNAHV ASCDAIMTRC LAVHECFVKR VDWSVEYPII

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1865       1875       1885       1895       1905       1915
EMCR        ANEKFINGCG RNVQGHVVRA ALKLYKPSVI HDIGNPKGVR CA-VTDAKWY CYDKQPVNSN
229E        ANENAINKGG RTVQSHIMRA AIKLYNPKAI HDIGNPKGIR CA-VTDAKWY CYDKNPINSN
PEDV        GNEAVINKSG RIVQSHTMRS VLKLYNPKAI YDIGNPKGIR CA-VTDAKWF CFDKNPTNSN
TGEV        DNEEKINKAG RIVQSHVMKA ALKIFNPAAI HDVGNPKGIR CA-TTPIPWF CYDRDPINNN
BoCoV       SNELSINTSC RVLQRVMLKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK FYDAQPIVKS
OC43        SNELSINTSC RVLQRVILKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK FYDAQPIVKS
MHV         SNEVSVNTSC RLLQRVMFRA AMLCNRYDVC YDIGNPKGLA CV--KGYDFK FYDASPVVKS
AIPV        ANEDEVNSSC RYLQRMYLNA CVDALKVNVV YDIGNPKGIK CVRRGDVNFR FYDKNPIVRN
SARS CoV    GDELRVNSAC RKVQHMVVKS ALLADKFPVL HDIGNPKAIK CVPQAEVEWK FYDAQPCSDK

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1925       1935       1945       1955       1965       1975
EMCR        ---VKLLDYD YATHG--QLD GLCLFWNCNV DMYPEFSIVC RFDTRTRSVF NLEGVNGGSL
229E        ---VKTLEYD YMTHG--QMD GLCLFWNCNV DMYPEFSIVC RFDTRTRSTL NLEGVNGGSL
PEDV        ---VKTLEYD YITHG--QFD GLCLFWNCNV DMYPEFSVVC RFDTRCRSPL NLEGCNGGSL
TGEV        ---VRCLDYD YMVHG--QMN GLMLFWNCNV DMYPEFSIVC RFDTRTRSKL SLEGCNGGAL
BoCoV       ---VKTLLYF FEAHKDSFKD GLCMFWNCNV DKYPPNAVVC RFDTRVLNNL NLPGCNGGSL
OC43        ---VKTLLYS FEAHKDSFKD GLCMFWNCNV DKYPPNAVVC RFDTRVLNNL NLPGCNGGSL
MHV         ---VKQFVYK YEAHKDQFLD GLCMFWNCNV DKYPANAVVC RFDTRVLNKL NLPGCNGGSL
AIPV        ---VKQFEYD YNQHKDKFAD GLCMFWNCNV DCYPDNSLVC RYDTRNLSVF NLPGCNGGSL
SARS CoV    AYKIEELFYS YATHHDKFTD GVCLFWNCNV DRYPANAIVC RFDTRVLSNL NLPGCDGGSL

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               1985       1995       2005       2015       2025       2035
EMCR        YVNKHAFHTP AYDKRAFVKL KPMPFFYFDD SDCDVVQ--- -EQVNYVPLR ASSCVTRCNI
229E        YVNNHAFHTP AYDKRAMAKL KPAPFFYYDD GSCEVVH--- -DQVNYVPLR ATNCITKCNI
PEDV        YVNNHAFHTP AFDKRAFAKL KPMPFFYYDD TECDKLQ--- -DSINYVPLR ASNCITKCNV
TGEV        YVNNHAFHTP AYDRRAFAKL KPMPFFYYDD SNCELVD--- -GQPNYVPLK SNVCITKCNI
BoCoV       YVNKHAFHTK PFSRAAFEHL KPMPFFYYSD TPCVYMDGMD AKQVDYVPLK SATCITRCNL
OC43        YVNKHAFHTK PFARAAFEHL KPMPFFYYSD TPCVYMDGMD AKQVDYVPLK SATCITRCNL
MHV         YVNKHAFHTS PFTRAAFENL KPMPFFYYSD TPCVYMEGME SKQVDYVPLR SATCITRCNL
AIPV        YVNKHAFYTP KFDRISFRNL KAMPFFFYDS SPCETIQ-VD GVAQDLVSLA TKDCITKCNI
SARS CoV    YVNKHAFHTP AFDKSAFTNL KQLPFFYYSD SPCESHGKQV VSDIDYVPLK SATCITRCNL

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2045       2055       2065       2075       2085       2095
EMCR        GGAVCSKHAN LYQKYVEAYN TFTQAGFNIW VPHSFDVYNL WQIFIET-NL QSLENIAFNV
229E        GGAVCSKHAN LYRAYVESYN IFTQAGFNIW VPTTFDCYNL WQTFTEV-NL QGLENIAFNV
PEDV        GGAVCSKHCA MYHSYVNAYN TFTSAGFTIW VPTSFDTYNL WQTFSN--NL QGLENIAFNV
TGEV        GGAVCKKHAA LYRAYVEDYN IFMQAGFTIW CPQNFDTYML WHGFVNSKAL QSLENVAFNV
BoCoV       GGAVCLKHAE EYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTK---L QSLENVVYNL
OC43        GGAVCLKHAE EYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTK---L QSLENVVYNL
MHV         GGAVCLKHAE DYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTR---L QSLENVVYNL
AIPV        GGAVCKKHAQ MYAEFVTSYN AAVTAGFTFW VTNKLNPYNL WKSFSA---L QSIDNIAYNM
SARS CoV    GGAVCRHHAN EYRQYLDAYN MMISAGFSLW IYKQFDTYNL WNTFTR---L QSLENVAYNV

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2105       2115       2125       2135       2145       2155
EMCR        VKKGCFTGVD GELPVAVVND KVFVRYGDVD NLVFTNKTTL PTNVAFELFA KRKMGLTPPL
229E        VNKGSFVGAD GELPVAISGD KVFVRDGNTD NLVFVNKTSL PTNIAFELFA KRKVGLTPPL
PEDV        LKKGSFVGDE GELPVAVVND KVLVRDGTVD TLVFTNKTSL PTNVAFELYA KRKVGLTPPI
TGEV        VKKGAFTGLK GDLPTAVIAD KIMVRDGPTD KCIFTNKTSL PTNVAFELYA KRKLGLTPPL
BoCoV       VKTGHYTGQA GEMPCAIIND KVVAKIDKED VVIFINNTTY PTNVAVELFA KRSIRHHPEL
OC43        VKTGHYTGQA GEMPCAIIND KVVAKIDKED VVIFINNTTY PTNVAVELFA KRSVRHHPEL
MHV         VNAGHFDGRA GELPCAVIGE KVIAKIQNED VVVFKNNTPF PTNVAVELFA KRSIRPHPEL
AIPV        YKGGHYDAIA GEMPTVITGD KVFVIDQGVE KAVFVNQTTL PTSVAFELYA KRNIRTLPNN
SARS CoV    VNKGHFDGHA GEAPVSIINN AVYTKVDGID VEIFENKTTL PVNVAFELWA KRNIKPVPEI

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               2165       2175       2185       2195       2205       2215
EMCR        SILKNLGVVA TYKFVLWDYE AERPFTSYTK SVCKYTDFN- -------EDV CVCFDNSIQG
```

```
                229E        SILKNLGVVA TYKFVLWDYE AERPLTSFTK SVCGYTDFA- -------EDV CTCYDNSIQG
                PEDV        TILRNLGVVC TSKCVIWDYE AERPLTTFTK DVCKYTDFE- -------GDV CTLFDNSIVG
                TGEV        TILRNLGVVA TYKFVLWDYE AERPFSNFTK QVCSYTDLD- -------SEV VTCFDNSIAG
                BoCoV       KLFRNLNIDV CWKHVIWDYA RESIFCSNTY GVCMYTDLK- -----LIDKL NVLFDGRDNG
          5     OC43        KLFRNLNIDV CWKHVIWDYA RESIFCSNTY GVCMYTDLK- -----FIDKL NVLFDGRDNG
                MHV         KLFRNLNIDV CWSHVLWDYA KDSVFCSSTY KVCKYTDLQ- -----CIESL NVLFDGRDNG
                AIPV        RILKGLGVDV TNGFVIWDYA NQTPLYRNTV KVCAYTDIE- ------PNGL VVLYDDR-YG
                SARS CoV    KILNNLGVDI AANTVIWDYK REAPAHVSTI GVCTMTDIAK KPTESACSSL TVLFDGRVEG

         10                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 2225       2235       2245       2255       2265       2275
                EMCR        SYERFTLTTN AVLFSTVVIK N----LTPIK LNFGMLNGMP VSSIKSDKGV EKLVNWYTYV
                229E        SYERFTLSTN AVLFSATAVK TGGKSLPAIK LNFGMLNGNA IATVKSEDGN IKNINWFVYV
                PEDV        SLERFSMTQN AVLMSLTAVK K----LTGIK LTYGYLNGVP VNTHED---- -KPFTWYIYT
         15     TGEV        SFERFTTTRD AVLISNNAVK G----LSAIK LQYGLLNDLP VSTVGN---- -KPVTWYIYV
                BoCoV       ALEAFKRSNN GVYISTTKVK S----LSMIR GPPRAELNGV VVDKVGD--- -TDCVFYFAV
                OC43        ALEAFKRSNN GVYISTTKVK S----LSMIR GPPRAELNGV VVDKVGD--- -TDCVFYFAV
                MHV         ALEAFKKCRD GVYINTTKIK S----LSMIK GPQRADLNGV VVEKVGD--- -SDVEFWFAM
                AIPV        DYQSFLAADN AVLVSTQCYK R----YSYVE IPSNLLVQNG MPLKDG---- ---ANLYVYK
         20     SARS CoV    QVDLFRNARN GVLITEGSVK G----LTPSK GPAQASVNGV TLIGES---- -VKTQFNYFK

                                 2285       2295       2305       2315       2325       2335
                EMCR        RKNG------ -----QFQDH Y--------- ---------- ----DGFYTQ GRNLSDFTPR
         25     229E        RKDG------ -----KPVDH Y--------- ---------- ----DGFYTQ GRNLQDFLPR
                PEDV        RKNG------ -----KFEDY P--------- ---------- ----DGYFTQ GRTTADFSPR
                TGEV        RKNG------ -----EYVEQ I--------- ---------- ----DSYYTQ GRTFETFKPR
                BoCoV       RKEGQDVIFS QFDSLRVSSN QSPQGNLGSN -EPGNVGGND ALATSTIFTQ SRVISSFTCR
                OC43        RKEGQDVIFS QFDSLGVSSN QSPQGNLGSN GKPGNVGGND ALSISTIFTQ SRVISSFTCR
         30     MHV         RRDGDDVIFS RTGSLEPSHY RSPQGNPGGN -RVGDLSGNE ALARGTIFTQ SRFLSSFAPR
                AIPV        RVNG------ -----AFVTL P--------- ---------- ----NTINTQ GRSYETFEPR
                SARS CoV    KVDG------ ----=IIQQL P--------- ---------- ----ETYFTQ SRDLEDFKPR

         35                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 2345       2355       2365       2375       2385       2395
                EMCR        SDMEYDFLNM DMGVFINKYG LEDFNFEHVV YGDVSKTTLG GLHLLISQFR LSKMGVLKAD
                229E        STMEEDFLNM DIGVFIQKYG LEDFNFEHVV YGDVSKTTLG GLHLLISQVR LSKMGILKAE
                PEDV        SDMEKDFLSM DMGLFINKYG LEDYGFEHVV YGDVSKTTLG GLHLLISQVR LACMGVLKID
                TGEV        STMEEDFLSM DTTLFIQKYG LEDYGFEHVV FGDVSKTTIG GMHLLISQVR LAKMGLFSVQ
         40     BoCoV       TDMEKDFIAL DQDVFIQKYG LEDYAFEHIV YGNFNQKIIG GLHLLIGLYR RQQTSNLVIQ
                OC43        TDMEKDFIAL DQDVFIQKYG LEDYAFEHIV YGNFNQKIIG GLHLLIGLYR RQQTSNLVVQ
                MHV         SEMEKDFMDL DEDVFIAKYS LQDYAFEHVV YGSFNQKIIG GLHLLIGLAR RQQKSNLVIQ
                AIPV        SDIERDFLAM SEESFVERYG -KDLGLQHIL YGEVDKPQLG GLHTVIGMYR LLRANKLNAK
                SARS CoV    SQMETDFLEL AMDEFIQRYK LEGYAFEHIV YGDFSHGQLG GLHLMIGLAK RSQDSPLKLE
         45
                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 2405       2415       2425       2435       2445       2455
                EMCR        DFVTASDTTL RCCTVTYLNE LSSKVVCTYM DLLLDDFVTI LK---SLDLG VISKVHEVII
                229E        EFVAASDITL KCCTVTYLND PSSKTVCTYM DLLLDDFVSV LK---SLDLT VVSKVHEVII
         50     PEDV        EFVSSNDSTL KSCTVTYADN PSSKMVCTYM DLLLDDFVSI LK---SLDLS VVSKVHEVMV
                TGEV        EFMNNSDSTL KSCCITYADD PSSKNVCTYM DILLDDFVTI IK---SLDLN VVSKVVDVIV
                BoCoV       EFVS-YDSSI HSYFITDEKS GGSKSVCTVI DILLDDFVAL VK---SLNLN CVSKVVNVNV
                OC43        EFVS-YDSSI HSYFITDEKS GGSKSVCTVI DILLDDFVAL VK---SLNLN CVSKVVNVNV
                MHV         EFVP-YDSSI HSYFITDENS GSSKSVCTVI DLLLDDFVDI VK---SLNLN CVSKVVNVNV
         55     AIPV        SVTN-SDSDV MQNYFVLSDN GSYKQVCTVV DLLLDDFLEL LRNILKEYGT NKSKVVTVSI
                SARS CoV    DFIP-MDSTV KNYFITDAQT GSSKCVCSVI DLLLDDFVEI IK---SQDLS VISKVVKVTI

                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 2465       2475       2485       2495       2505       2515
         60     EMCR        DNKPYRWMLW CKDNHLSTFY PQLQS-AEWK CGYAMPQIYK LQRMCLEPCN LYNYGAGIKL
                229E        DNKPWRWMLW CKDNAVATFY PQLQS-AEWK CGYSMPGIYK TQRMCLEPCN LYNYGAGLKL
                PEDV        DCKMWRWMLW CKDHKLQTFY PQLQA-SEWK CGYSMPSIYK IQRMCLEPCN LYNYGAGVKL
                TGEV        DCKAWRWMLW CENSHIKTFY PQLQS-AEWN PGYSMPTLYK IQRMCLERCN LYNYGAQVKL
                BoCoV       DFKDFQFMLW CNDEKVMTFY PRLQAASDWK PGYSMPVLYK YLNSPMERVS LWNYGKPVTL
         65     OC43        DFKDFQFMLW CNDEKVMTFY PRLQAASDWK PGYSMPVLYK YLNSPMERVS LWNYGKPVTL
                MHV         DFKDFQFMLW CNEEKVMTFY PRLQAAADWK PGYVMPVLYK YLESPLERVN LWNYGKPITL
                AIPV        DYHSINFMTW FEDGSIKTCY PQLQS--AWT CGYNMPELYK VQNCVMEPCN IPNYGVGITL
                SARS CoV    DYAEISFMLW CKDGHVETFY PKLQASQAWQ PGVAMPNLYK MQRMLLEKCD LQNYGENAVI

         70                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                 2525       2535       2545       2555       2565       2575
                EMCR        PSGIMLNVVK YTQLCQYLNS TTMCVPHNMR VLHYGAGSDK GVAPGTTVLK RWLPPD----
                229E        PSGIMFNVVK YTQLCQYFNS TTLCVPHNMR VLHLGAGSDY GVAPGTAVLK RWLPHD----
                PEDV        PDGIMFNVVK YTQLCQYLNS TTMCVPHHMR VLHLGAGSDK GVAPGTAVLR RWLPLD----
         75     TGEV        PDGITTNVVK YTQLCQYLNT TTLCVPHKMR VLHLGAAGAS GVAPGSTVLR RWLPDD----
                BoCoV       PTGCMMNVAK YTQLCQYLNT TTLAVPVNTR VLHLGAGSEK GVAPGSAVLR QWLPAGTILR
                OC43        PTGCMMNVAK YTQLCQYLNT TTLAVPVNMR VLHLGAGSEK GVAPGSAVLR QWLPAG----
                MHV         PTGCLMNVAK YTQLCQYLNT TTLAVPANMR VLHLGAGSDK DVAPGSAVLR QWLPAG----
                AIPV        PSGILMNVAK YTQLCQYLSK TTICVPHNMR VMHFGAGSDK GVAPGSTVLK QWLPEG----
         80     SARS CoV    PKGIMMNVAK YTQLCQYLNT LTLAVPYNMR VIHFGAGSDK GVAPGTAVLR QWLPTG----
```

```
                    2585       2595       2605       2615       2625       2635
EMCR        ------AIII DNDINDYVSD ADFSITGDCA TVYLEDKFDL LISDMYDG-- --RIKFCDGE
229E        ------AIVV DNDVVDYVSD ADFSVTGDCA TVYLEDKFDL LISDMYDG-- --RTKAIDGE
PEDV        ------AIIV DNDSVDYVSD ADYSVTGDCS TLYLSDKFDL VISDMYDG-- --KIKSCDGE
TGEV        ------AILV DNDLRDYVSD ADFSVTGDCT SLYIEDKFDL LVSDLYDG-- --STKSIDGE
BoCoV       QWLPAGTILV HNDLYPFVSD SVATYFGDCI TLPFDCQWDL IISDMYD--- --LLLDIGVH
OC43        ------TILV DNDLYPFVSD SVATYFGDCI TLPFDCQWDL IISDMYDP-- --ITKNIGEY
MHV         ------SILV DNDINPFVSD SVASYYGNCI TLPIACQWDL IISDMYDP-- --LTKNIGEY
AIPV        ------TLLV DNDIVDYVSD AHVSVLSDCN KYNTEHKFDL VISDMYTDND SKRKHEGVIA
SARS CoV    ------TLLV DSDLNDFVSD ADSTLIGDCA TVHTANKWDL IISDMYDP-- --RTKHVTKE

                    2645       2655       2665       2675       2685       2695
EMCR        NVSKDGFFTY LNGVIREKLA IGGSVAIKIT EYSWNKYLYE LIQRFAFWTL FCTSVNTSSS
229E        NVSKEGFFTY INGFICEKLA IGGSIAIKVT EYSWNKKLYE LVQRFSFWTM FCTSVNTSSS
PEDV        NVSKEGFFPY INGVITEKLA LGGTVAIKVT EFSWNKKLYE LIQKFEYWTM FCTSVNTSSS
TGEV        NTSKDGFFTY INGFIKEKLS LGGSVAIKIT EFSWNKDLYE LIQRFEYWTV FCTSVNTSSS
BoCoV       VVRCS---YI HCHMIRDKLA LGGSVAIKIT EFSWNAELYK LMGYFAFWTV FCTNANASSS
OC43        NVSKDGFFTY ICHMIRDKLA LGGSVAIKIT EFSWNAELYK LMGYFAFWTV FCTNANASSS
MHV         NVSKDGFFTY LCHLIRDKLA LGGSVAIKIT EFSWNAELYS LMGKFAFWTI FCTNVNASSS
AIPV        NNGNDDVFIY LSSFLRNNLA LGGSFAVKVT ETSWHEVLYD IAQDCAWWTM FCTAVNASSS
SARS CoV    NDSKEGFFTY LCGFIKQKLA LGGSIAVKIT EHSWNADLYK LMGHFSWWTA FVTNVNASSS

                    2705       2715       2725       2735       2745       2755
EMCR        EAFLIGINYL GDFIQGPFIA GNTVHANYIF WRNSTIMSLS YNSVLDLSKF ECKHKATVVV
229E        EAFVVGINYL GDFAQGPFID GNIIHANYVF WRNSTVMSLS YNSVLDLSKF NCKHKATVVV
PEDV        EAFLIGVHYL GDFASGAVID GNTMHANYIF WRNSTIMTMS YNSVLDLSKF NCKHKATVVV
TGEV        EGFLIGINYL GPYCDKAIVD GNIMHANYIF WRNSTIMALS HNSVLDTPKF KCRCNNALIV
BoCoV       EGFLIGINYL GK--PKVEID GNVMHAIICF G--------- ---------- EIPQFGTGVL
OC43        EGFLIGINYL CK--PKVEID GNVMHANYLF WRNSTVWNGG AYSLFDMAKF PLKLAGTAVI
MHV         EGFLIGINWL NR--TRTEID GKTMHANYLF WRNSTMWNGG AYSLFDMSKF PLKVAGTAVV
AIPV        EAFLIGVNYL GAS-EKVKVS GKTLHANYIF WRNCNYLQTS AYSIFDVAKF DLRLKATPVV
SARS CoV    EAFLIGANYL GK--PKEQID GYTMHANYIF WRNTNPIQLS SYSLFDMSKF PLKLRGTAVM

                    2765       2775       2785       2795
EMCR        TLKDSDVNDM VLSLIKSGRL LLRNSGRFGG FSNHLVSTK-
229E        QLKDSDINEM VLSLVRSGKL LVRGNGKCLS FSNHLVSTK-
PEDV        NLKDSSISDV VLGLLKNGKL LVRNNDAICG FSNHLVNVNK
TGEV        NLKEKELNEM VIGLLRKGKL LIRNNGKLLN FGNHFVNTP-
BoCoV       IACLIWLNSR LSWLVMP--- ---------- ----------
OC43        NLRADQINDM VYSLLEKGKL LIRDTNKEVF VGDSLVNVI-
MHV         SLKPDQINDL VLSLIEKGKL LVRDTRKEVF VGDSLVNVK-
AIPV        NLKTEQKTDL VFNLIKCGKL LVRDVGNTSF TSDSFVCTM-
SARS CoV    SLKENQINDM IYSLLEKGRL IIRENNRVVV SSDILVNN--
```

d. Putative Orf 1ab

```
                      5         15         25         35         45         55
EMCR        ---------- ---------M FYNQVTLAVA SDSEISGFGF AIPSVAVRAY SEAAAQGFQA
229E        ---------- ---------M ACNRVTLAVA SDSEISANGC STIAQAVRRY SEAASNGFRA
PEDV        ---------- ---------M ASNHVTLAFA NDAEISAFGF CTASEAVSYY SEAAASGFMQ
TGEV        ---------- ---------M SSKQFKILVN EDYQVNVPSL PIR-DVLQEI KYCYRNGFEG
OV43        MSKINKYGLE LHWAPEFPWM FEDAEEKLDN PSSSEVDMIC STTAQKLETD GICPENHVMV
BoCoV       MSKINKYGLE LHWAPEFPWM FEDAEEKLDN PSSSEVDIVC STTAQKLETG GICPENHVMV
MHV         MAKMGKYGLG FKWAPEFPWM LPNASEKLGS PERSEEDGFC PSAAQEPKTK GKTLINHVRV
AIBV        ---------- ---------- MASSLKQGVS PKPRDVILVS KDIPEQLCDA LFFYTSHNPK
SARS CoV    ---------- --MESLVLGV NEKTHVQLSL PVLQVRDVLV RGFGDSVEEA LSEAREHLKN

                      65         75         85         95        105        115
EMCR        CRFVAFGLQD CVTGINDDDY VIALTG---- ---------- TNQLCAKILL FSDRPLNLRG
229E        CRFVSLDLQD CIVGIADDTY VMGLHG---- ---------- NQTLFCNIMK FSDRPFMLHG
PEDV        CRFVSLDLAD TVEGLLPEDY VMVVIG---- ---------- TTKLSAYVDT FGSRPRNICG
TGEV        YVFVPEYCRD LVDCDRKDHY VIGVLG---- ---------- NGVSDLKPVL LTEPSVMLQG
OV43        DCRRLLKQEC CVQSSLIREI VMNASPYDLE VLLQDALQSR EAVLVTTPLG MSLEACYVRG
BoCoV       DCRRLLKQEC CVQSSLIREI VMNTRPYDLE VLLQDALQSR EAVLVTPPLG MSLEACYVRG
MHV         DCSRLPALEC CVQSAIIRDI FVDEDPLNVE ASTMMALQFG SAVLVKPSKR LSIQAWAKLG
AIBV        DYADAFAVRQ KFDRSLQTGK QFKFET---- ---------V CGLFLLKGVD KITPGVPAKV
SARS CoV    GTCGLVELEK GVLPQLEQPY VFIKR--SDA LSTNHGHKVV ELVAEMDGIQ YGRSGITLGV

                     125        135        145        155        165        175
```

```
EMCR       WLIFSNSNYV LQDFDVVFG- -HGAGSVVFV DKYMCGFDGK PVLP--KNMW EFRDYFNDNT
229E       WLVFSNSNYL LEEFDVVFGK -RGGGNVTYT DQYLCGADGK PVMS--EDLW QFVDHFGENE
PEDV       WLLFSNCNYF LEELELTFG- -RRGGNIVPV DQYMCGADGK PVLQ--ESEW EYTDFFADSE
TGEV       FIVRANCNGV LEDFDLKIA- -RTGRGAIYV DQYMCGADGK PVIE--G--- DFKDYFGDED
OV43       CNPKGWTMGL FRRRSVCNTG RCTVNKHVAY QLYMIDPAGV CLGAGQFVGW VIPLAFMPVQ
BoCoV      CNPNGWTMGL FRRRSVCNTG RCAVNKHVAY QLYMIDPAGV CFGAGQFVGW VIPLAFMPVQ
MHV        VLPKTPAMGL FKRFCLCNTR ECVCDAHVAF QLFTVQPDGV CLGNGRFIGW FVPVTAIPAY
AIBV       LKATSKLADL EDIFGVSPLA RKYRELLKTA CQWSLTVEAL DVR-----AQ TLDEIFDPTE
SARS CoV   LVPHVGETPI AYRNVLLRK- --NGNKGAGG HSYGIDLKSY DLG--DELGT DPIEDYEQNW

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              185        195        205        215        225        235
EMCR       DS-IVIGGVT YQLAWDVIRK DLSYEQQNVL AIESIHYLG- TTGHTLKSGC KLINAKPPKY
229E       E--IIINGHT YVCAWLTKRK PLDYKRQNNL AIEEIEYVHG DALHTLRNGS VLEMAKEVKT
PEDV       DGQLNIAGIT YVKAWIVERS DVSYASQNLT SIKSITYCS- TYEHTFLDGT AMKVARTPKI
TGEV       --IIEFEGEE YHCAWTTVRD EKPLNQQTLF TIQEIQYNL- DIPHKLPNCA TRHVAPPVKK
OV43       SRKFIVPWVM YLRKRGEKGA YNKDHGRGGF GH-VYDFKVE DAYDQVHDEP KGKFSKKAYA
BoCoV      SRKFIAPWVM YLRKCGEKGA YIKDYKRGGF EH-VYNFKVE DAYDLVHDEP KGKFSKKAYA
MHV        AKQWLQPWSI LLRKGGNKGS VTSGHFRRAV TMPVYDFNVE DACEEVHLNP KGKYSRKAYA
AIBV       ILWLQVAAKI HVSSMAMRRL VGEVTAKVMD ALG------- SNLSALFQIV KQQIARIFQK
SARS CoV   NTKHGSGALR ELTRELNGGA VTRYVDNNFC GPDGYPLDCI KDFLARAGKS MCTLS-EQLD

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              245        255        265        275        285        295
EMCR       SSKVVLSGEW NAVYKAFGSP FITNGISLLD IIVKPVFFNA FVKCNCGSEN WSVGAWDGYL
229E       SSKVVLSDAL DKLYKVFGSP VMTNGSNILE AFTKPVFISA LVQCTCGTKS WSVGDWTGFK
PEDV       KKNVVLSEPL ATIYREIGSP FVDNGSDARS IIRRPVFLHA FVKCKCGSYH WTVGDWTSYV
TGEV       NSKIVLSEDY KKLYDIFGSP FMGNGDCLSK CFDTLHFIAA TLRCPCGSES SGVGDWTGFK
OV43       LIRGYRGVKP LLYVDQYGCD YTGSLADGLE AYADKTLQEM KALFPTWSQE LLFDVIVAWH
BoCoV      LIRGYRGVKP LLYVDQYGCD YTGGLADGLE AYADKTLQEM KALFPIWSQE LPFDVTVAWH
MHV        LLKGYRGVKS ILFLDQYGCD YTGRLAKGLE DYGDCTLEEM KELFPVWCDS LDNEVVVAWH
AIBV       ALAIFENVNE LPQRIAALKM AFAKCARSIT VVVVERTLVV KEFAGTCLAS INGAVAKFFE
SARS CoV   YIESKRGVYC CRDHEHEIAW FTERSDKSYE HQTPFEIKSA KKFDTFKGEC PKFVFPLNSK

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              305        315        325        335        345        355
EMCR       SSCCGTPAKK LCVVPGNVVP GDVIITSTDA GCGVKYYAGL VVKHITNITG VSLWRVTAVH
229E       SSCCNVISNK LCVVPGNVKP GDAVITTQQA GAGIKYFCGM TLKFVANIEG VSVWRVIALQ
PEDV       STCCGFKCKP VLVASCSAMP GSVVVTRAGA GTGVKYYNNM FLRHVADIDG LAFWRILKVQ
TGEV       TACCGLSGKV KGVTLGDIKP GDAVVTSMSA GKGVKFFANC VLQYAGDVEG VSIWKVIKTF
OV43       VVRDPRYVMR LQSAATIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQYNLVDIM
BoCoV      VVRDPRYVMR LQSASTIRSV AYVANPTEDL CDGSVVIKEP VHVYADDSII LRQHNLVDIM
MHV        VDRDPRAVMR LQTLATIRSI GYVGQPTEDL VDGDVVVREP AHLLAANAIV KRLPRLVETM
AIBV       ELPNGFMGSK IFTTLAFFKE AAVRVVENIP NAPRGTKGFE VVGNAKGTQV VVRGMRNDLT
SARS CoV   VKVIQPRVEK KKTEGFMGRI RSVYPVASPQ ECNNMHLSTL MKCNHCDEVS WQTCDFLKAT

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              365        375        385        395        405        415
EMCR       SDGMFVATSS YDALLHRNSL DPFCFDVNTL LSNQLRLAFL GASVTEDVKF A------AST
229E       SVDCFVASST FVEEEHVNRM DTFCFNVRNS VTDECRLAML GAEMTSNVRR Q------VAS
PEDV       SKDDLACSGK FLEHHEEGFT DPCYFLNDSS LATKLKFDIL SGKFSDEVKQ A------IIA
TGEV       TVDETVCTPG FEGELN---- DFIKPESKSL VACSVKRAFI TGDIDDAVHD C------IIT
OV43       SHFYMEADTV VNAFYGVALK DCGFVMQFGY IDCEQDSCDF KGWIPGNMID G---FACTTC
BoCoV      SCFYMEADAV VNAFYGVDLK DCGFVMQFGY IDCEQDLCDF KGWVPGNMID G---FACTTC
MHV        LYT---DSSV TEFCYKTKLC DCGFITQFGY VDCCGDACDF RGWVPGNMMD G---FLCPGC
AIBV       LLDQKADIPV EPEGWS---- --AILDGHLC YVFRSGDRFY AAPLSGNFAL S---------
SARS CoV   CEHCGTENLV IEGPTTCGYL PTNAVVKMPC PACQDPEIGP EHSVADYHNH SNIETRLRKG

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              425        435        445        455        465        475
EMCR       GVIDISAGMF GLYDDILTNN KPWFVRKASG LFDAIWDAFV AAIKLVPTTT GGLVRFVKSI
229E       GVIDISTGWF DVYDDIFAES KPWFVRKAED IFGPCWSALA SALKQLKVTT GELVRFVKSI
PEDV       GHVVVGSALV DIVDDALG-- QPWFIRKLGD LASAPWEQLK AVVRGLGLLS DEVVLFGKRL
TGEV       GKLDLSTNLF GNVGLLFKK- TPWFVQKCGA LFVDAWKVVE ELCGSLTLTY KQIYEVVASL
OV43       GHVYEVGDLI AQSSGVLPVN PVLHTKSAAG YGGFGCKDSF TLYGQTVVYF GGCVYWSPAR
BoCoV      GHVYETGDLL AQSSGVLPVN PVLHTKSAAG YGGFGCKDSF TLYGQTVVYF GGCVYWSPAR
MHV        SKSYMPWELE AQSSGVIPKG GVLFTQSTDT VN----RESF KLYGHAVVPF GSAVYWSPYP
AIBV       -DVHCCERVV CLSDGVTP-- --EIN--DGL ILAAIYSSFS VSELVTALKK GEPFKFLGHK
SARS CoV   GRTRCFGGCV FAYVGCYNKR AYWVPRASAD IG----SGHT GITGDNVETL NEDLLEILSR

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              485        495        505        515        525        535
EMCR       ASTVLTVSNG VIIMCADVPD AFQPVYRTFT QAICAAFDFS LDVFKIG--- ----------
229E       CNSAVAVVGG TIQILASVPE KFLNAFDVFV TAIQTVFDCA VETCTIA--- ----------
PEDV       SCATLSIVNG VFEFLADVPE KLAAAVTVFV NFLNEFFESA CDCLKVG--- ----------
TGEV       CTSAFTIVNY KPTFVVPD-N RVKDLVDKCV KVLVKAFDVF TQIITIAG-- ----------
OV43       NIWIPILKSS VKSYDSLVYT GVLGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS
BoCoV      NIWIPILKSS VKSYDGLVYT GVVGCKAIVK ETNLICKALY LDYVQHKCGN LHQRELLGVS
MHV        GMWLPVIWSS VKSYADLTYT GVVGCKAIVQ ETDAICRSLY MDYVQHKCGN LEQRAILGLD
AIBV       ---FVYAKDA AVSFTLAKAA TIADVLRLFQ SARVIAEDVW SSFTEKS--- ----------
```

```
                SARS CoV    ---ERVNINI VGDFHLNEEV AIILASFSAS TSAFIDTIKS LDYKSFKT-I VESCGNYKVT

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              545        555        565        575        585        595
                EMCR        ----DVKFKR LGDYVLTENA LVRLTTEVVR GVRD------ ---------- ----A-----
                229E        ----GKAFDK VFDYVLLDNA LVKLVTTKLK GVRE------ ---------- ----R-----
                PEDV        ----GKTFNK VGSYVLFDNA LVKLVKAKAR GPRQ------ ---------- ----A-----
                TGEV        --IEAKCFVL GAKYLLFNNA LVKLVSVKIL GKKQ------ ---------- ----K-----
                OV43        DVWHKQLLLN RGVYKPLLEN IDYFNMRRAK FSLETFT--- VCADGFMPFL LDDLVPRAYY
                BoCoV       DVWHKQLLLN RGVYKPLLEN IDYFNMRRAK FSLETFT--- VCADGFMPFL LDDLVPRAYY
                MHV         DVYHRQLLVN RGDYSLLLEN VDLFVKRRAE FACK-FA--- TCGDGLVPLL LDGLVPRSYY
                AIBV        ---------- FEFWKLAYGK VRNLEEFVKT YVCK------ ---------- ----------
                SARS CoV    KGKPVKGAWN IGQQRSVLTP LCGFPSQAAG VIRSIFARTL DAANHSIPDL QRAAVTILDG

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              605        615        625        635        645        655
                EMCR        -RIKKAMFTK VVVGPTTEVK FSVIELATVN LRLVDCAPVV CPKGKIVVIA GQAFFYSGGF
                229E        -GLNKVKYAT VVVGSTEEVK SSRVERSTAV LTIANNYSKL FDEGYTVVIG DVAYFVSDGY
                PEDV        -GICEVRYTS LVVGSTTKVV SKRVENANVN LVVVDEDVTL NTTGRTVVVD GLAFFESDGF
                TGEV        -GLECAFFAT SLVGATVNVT PKRTETATIS LNKVDDVVAP G-EGYIVIVG DMAFYKSGEY
                OV43        LAVSGQAFCD YADKLCHAVV SKSKELLDVS LDSLGAAIHY LNSKIVDLAQ HFSDFGTSFV
                BoCoV       LAVSGQAFCD YAGKICHAVV SKSKELLDVS VDSLGAAIHY LNSKIVDLAQ HFSDFGTSFV
                MHV         LIKSGQAFTS MMVNFSHEVT DMCMDMALLF MHDVKVATKY VKKVTGKLAV RFKALGVAVV
                AIBV        AQMSIVILAA VLGEDIWHLV SQVIYKLGVL FTKVVDFCDK HWKGFCVQLK RAKLIVTETF
                SARS CoV    ISEQSLRLVD AMVYTSDLLT NSVIIMAYVT GGLVQQTSQW LSNLLGTTVE KLRPIFEWIE

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              665        675        685        695        705        715
                EMCR        YRFMVDSTTV LNDPVFTGEL FYTIKFSGFK LDGFN----- ---HQFVNAS SATDAIIAVE
                229E        FRLMASPNSV LTTAVYKPLF AFNVNVMGTR PE-------- ----KFPTTV TCENLESAVL
                PEDV        YRHLADADVV IEHPVYKSAC ELKPVFECDP IP--D----- ---FPLPVAA SVAELCVQTD
                TGEV        YFMMSSPNFV LTNNVFKAVK VPSYDIVYDV DNDTKSKMIA KLGSSFEYDG DIDAAIVKVN
                OV43        SKIVHFFKTF TTSTALAFAW VLFHVLHGAY IVVESDIYFV KN-IPRYASA VAQAFQSVAK
                BoCoV       SKIVHFFKTF TTSTALAFAW VLFHVLHGAY IVVESDIYFG KN-IPRYASA VAQAFRSGAK
                MHV         RKITEWFDLA VDTAASAAGW LCYQLVNGLF AVANGGITFL SD-VPELVKN FVDKFKVFFK
                AIBV        CVLKGVAQHC FQLLLDAIHS LYKSFKKCAL GRIHG----- --DLLFWKGG VHKIVQDGDE
                SARS CoV    AKLSAGVEFL KD-----AWE ILKFLITGVF DIVKGQIQVA SDNIKDCVKC FIDVVNKALE

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              725        735        745        755        765        775
                EMCR        LLLSDFKTAV FVYTCVVDGC SVIVRRDAT- FATHVCFKDC YSIWEQFCID NCGE------
                229E        FVNDKITEFQ LDYSIDVIDN EIIVKPNIS- LCVPLYVRDY VDKWDDFCRQ YSNE------
                PEDV        LLLKNYNTPY KTYSCVVRGD KCCITCTLQ- FKAPSYVEDA VN-FVDLCTK NIGT------
                TGEV        ELLIEFRQQS LCFRAFKDDK SIFVEAYFKK YKMPACLAKH IG-LWNIIKK DSCK------
                OV43        VVLDSLRVTF IDGLSCFKIG RRRICLSGRK IYEVERG-LL HSSQLPLDVY DLTMPSQVQK
                BoCoV       VGLDSLRVTF IDGLSCFKIG RRRICLSGSK IYEVERG-LL HSSQLPLDVY DLTMPSQVQK
                MHV         VLIDSMSVSV LSGLTVVKTA SNRVCLAGCK VYEVVQK-RL SAYVMPVGCN EATC------
                AIBV        IWFDAIDSVD VEDLGVVQEK SIDFEVCDDV TLPENQPGHM VQIEDDGKNY MFFR------
                SARS CoV    MCIDQVTIAG AK-LRSLNLG EVFIAQSKGL YRQCIRGKEQ LQLLMPLKAP KEVT------

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              785        795        805        815        825        835
                EMCR        --------PW FLTDYNAILQ SNNPQCAIVQ ASESK----- --VLLERFLP KCPEILLSID
                229E        --------SW FEDDYRAFIS VLDITDAAVK AAESK----- --AFVDTIVP PCPSILKVID
                PEDV        --------AG FHEFYITAHE QQDLQGFLTT CCTMSG----- F-ECFMPTIP QCPAVLEEID
                TGEV        --------RG FLNLFNHLNE LEDIKETNIQ AIKN------ --------I- LCPDPLLDLD
                OV43        AKQKPIYLKG SGSDFSLADS VVEVVTTSLT PCG-----YS EPPKVADKIC IVDNVYMAKA
                BoCoV       TKQKGIYLKG SGSDFSLADS VVEVVTTSLT PCG-----YS EPPKVADKIC IVDNVYMAKA
                MHV         -------LVG EIEPAVVEDD VVDVVKAPLT YQG-----CC KPPTSFEKIC VVDKLYMAKC
                AIBV        -------FKK DENIYYTPMS QLGAINVVCK AGG------- ------KTVT FGETTVQEIP
                SARS CoV    ------FLEG DSHDTVLTSE EVVLKNGELE ALETPVDSFT NGAIVGTPVC VNGLMLLEIK

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              845        855        865        875        885        895
                EMCR        DGHLWNLFVE K--------- -FNFVTDWLK TLKLTLTSNG LLGNCAKRFR RVLVKLLDVY
                229E        GGKIWNGVIK N--------- -VNSVRDWLK SLKLNLTQQG LLGTCAKRFK RWLGILLEAY
                PEDV        GGSIWRSFIT G--------- -LNTMWDFCK RLKVSFGLDG IVVTVARKFK RLGALLAEMY
                TGEV        YGAIWYNCMP G--------- -CSDP-SVLG SVQLLIGNG- -VKVVCDGCK GFANQLSKGY
                OV43        GDKYYPVVVD -DHVGLLDQA WRVPCAG--R RVTFKEQPTV KEIISMPKII KVFYELDNDF
                BoCoV       GDKYYPVVVD -GHVGLLDQA WRVPCAG--R CVTFKEQPTV NEIASTPKTI KVFYELDKDF
                MHV         GDQFYPVVVD NDTIGVLDQC WRFPCAG--K KVEFNDKPKV KEIPST-RKI KINFALDATF
                AIBV        PPDVVPIKVS ---------- --IECCG--E PWNTIFKKAY KEPIEVDTDL TVEQLLSVIY
                SARS CoV    DKEQYCALSP --GLLATNNV FRLKGGAPIK GVTFG-EDTV WEVQGY-KNV RITFELDERV

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              905        915        925        935        945        955
                EMCR        NGFLETVCSV VHTAGVCIKY YAVNVP-YVV ISGFVSRVIR RERCD--VTF PCVSCVTFFY
                229E        NAFLDTVVST VKIGGLTFKT YAFDKP-YIV IRDIVCKVEN KTEAEWIELF PHNDRIKSFS
                PEDV        NTYLSTVVEN LVLAGVSFKY YATSVP-KIV LGGCFHSVKS VFASV--FQI PVQAGIEKFK
                TGEV        NKLCNAARND IEIGGIPFST FKTPTNTFIE MTDAIYSVIE QGKALS---- ----------
```

```
                OV43        NTILNTACGV FEVDDTVDME EFYAVVIDAI EEKLSPCKEL EGVGAKVSAF LQKLEDNPLF
                BoCoV       NTILNTACGE FEVDDTVDME EFYAVVIDAI EEKLSPCKEL EGVGAKVSAF LQKLEDNSLF
                MHV         DSVLSKACSE FEVDKDVTLD ELLDVVLDAV ESTLSPCKEH DVIGTKVCAL LNRLAEDYVY
                AIBV        EKMCDDLKLF PEAPEPPPFE NVALVDKNGK DLDCIKSCHL IYR------- ----------
                SARS CoV    DKVLNEKCSV YTVESGTEVT EFACVVAEAV VKTLQPVSDL LTN---MGID LDEWSVATFY

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  965        975        985        995       1005       1015
                EMCR        EFLDTCFGVS K-----PNAID VEHLELKETV FVEPKDGGQF FVSDDYLWYV V-DDIYYPAS
                229E        TFESAYMPIA D-----PTHFD IEEVELLDAE FVEPGCGGIL AVIDEHVFYK K-DGVYYPSN
                PEDV        VFLNCVHPVV -----PRVIE TSFVELEETT FKPPALNGGI AIVDGFAFYY D-GTLYYPTD
                TGEV        -FRDADVPVV DNGTISTADW SEPILLEPAE YVKPKNNGNV IVIAGYTFYK DEDEHFYPYG
                OV43        LFDEAGEEVL APKLYCAFTA P--EDDDFLE ESDVEEDDVE GEETDLTVTS AGQPCVASEQ
                BoCoV       LFDEAGEEVL APKLYCAFTA P--EDDDFLE ESGVEEDDVE GEETDLTVTS AGEPCVASEQ
                MHV         LFDEGGEEVI APKMYCSFSA P--DDEDCVA ADVVDADENQ GDDADDSAAL VTDTQEEDGV
                AIBV        -DYESDDDIE E------ED- AEECDTDSGE AEECDTNSEC EEEDEDTKVL ALIQDPASIK
                SARS CoV    LFDDAGEENF SSRMYCSFYP PDEEEEDDAE CEEEEIDETC EHEYGTEDDY QGLPLEFGAS

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1025       1035       1045       1055       1065       1075
                EMCR        CNGVLPVAFT KLAGGK--IS FSDDVIVHDV EPTHKVKLIF EFE------- -----DDVVT
                229E        GTNILPVAFT KAAGGK--VS FSDDVEVKDI EPVYRVKLCF EFE------- -----DEKLV
                PEDV        GNSVVPICFK KKGGGD--VK FSDEVSVKTI DPVYKVSLEF EFE------- -----SETIM
                TGEV        FGKIVQRMYN KMGGGDKTVS FSEEVDVQEI APVTRVKLEF EFD------- -----NEIVT
                OV43        EESSEVLEDT LDDGPSVETS DSQVEEDVEM SDFVDLESVI QD-------- ----YENVCF
                BoCoV       EESSEILEDT LDDGPCVETS DSQVEEDVQM SDFGDLESVI QD-------- ----YENVCF
                MHV         AKGQVGVAES DARLDQVEAF DIEKVEDPIL NELSAELNAP ADKTYEDVLA FDAIYSEALS
                AIBV        YPLPLDEDYS VYNGCIVHKD ALDVVNLPSG EETFVVNNCF EG-------- -------AVK
                SARS CoV    AETVRVEEEE EEDWLDDTTE QSEIEPEPEP TPEEPVNQFT GYLK------ ---LTDNVAI

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1085       1095       1105       1115       1125       1135
                EMCR        SLCKK----- ----------S FGKSIIYTG- DWEGLHEVLT SAMNVIG--- --QHIKLPQF
                229E        DVCEK----- ----------A IGKKIKHEG- DWDSFCKTIQ SALSVVS--- --CYVNLPTY
                PEDV        AVLNK----- ----------A VGNRIKVTG- GWDDVVEYIN VAIEVLK--- --DHVEVPKY
                TGEV        GVLER----- ----------A IGTRYKFTGT TWEEFEESIS EELDAIFDTL ANQGVELEGY
                OV43        EFYTT--EPE FVKVLGLYVP K--ATRNNCW LRSVLAVMQK LPCQFKDKNL QDLWVLYKQQ
                BoCoV       EFYTT--EPE FVKVLDLYVP K--ATRNNCW LRSVLAVMQK LPCQFKDKNL QDLWVLYKQQ
                MHV         AFYAVPGDET HFKVCGFYSP A--IERTNCW LRSTLIVMQS LPLEFKDLEM QKLWLSYKSS
                AIBV        PLPQK----- ---------- ----VVDVLG DWGEAVDAQE QLCQQEP--- ----LQHTFE
                SARS CoV    KCVDIVKEAQ SANPMVIVNA ANIHLKHGGG VAGALNKATN GAMQKESDDY IKLNGPLTVG

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1145       1155       1165       1175       1185       1195
                EMCR        YIYDEEGGYD VSKP--VMIS QWPISDDSDG CVVEASTDFH Q--LESVREE VD--------
                229E        YIYDEEGGND LSLP--VMIS EWPLSVQQAQ QEATLPDIAE D--VVDQVEE VNS-------
                PEDV        YIYDEEGGTD PNLP--VMVS QWPLNDDTIS QDLLDVEVVT DAPIDSEGDE VDSSAPEKVA
                TGEV        FIYDTCGGFD IKNPDGIMIS QYDINITADE KSEVSASSEE EE-VESVEED PENEIVEASE
                OV43        YSQLFVDTLV NKIPANIVLP QGGYVADFAY WFLTLCDWQC VAYWKCIKCD LALK---LKG
                BoCoV       YSQLFVDTLV NKIPANIVVP QGGYVADFAY WFLTLCDWQC VAYWKCIKCD LALK---LKG
                MHV         YNKEFVDKLV KSVPKSIILP QGGYVADFAY FFLSQCSFKA YANWRCLKCD MDLK---LQG
                AIBV        EPVENSTGSS KTMTEQVVVE DQELPVVEQD QDVVVYTPTD LEVAKETAEE VD--------
                SARS CoV    GSCLLSGHNL AKKCLHVVGP NLNAGEDIQL LKAAYENFNS QDILLAPLLS AGIFG--AKP

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1205       1215       1225       1235       1245       1255
                EMCR        ---------- ---------- -------IIE QPFGEVEHAL SIRQ------ ---PFSFSFR
                229E        ---------- ---------- -------IFD IETVDVKHDV S--------- ---PFEMPFE
                PEDV        ----------D VANSEPGDDG LPVAPETNVE SEVEEVAATL SFIKDTPSTV TKDPFAFDFV
                TGEV        GAEGTSSQEE VETVEVADIT STEEDVDIVE VSAKDDPWAA AVDVQEAEQF NPSLPPFKTT
                OV43        LDAMFFYGDV VSHICKCGES MVLIDVDVPF TAHFALKDKL FCAFITKRIV YKAACVVDVN
                BoCoV       LDAMFFYGDV VSHVCKCGES MVLIDVDVPF TAHFALKDKL FCAFITKRSV YKAACVVDVN
                MHV         LDAMFFYGDV VSHVCKCGTG MTLLSADIPY TLHFGLRDDK FCAFYTPRKV FRAACVVDVN
                AIBV        ---------- ---------- ------EFIL IFAVPKEEVV S--------- -----QKDGA
                SARS CoV    LQSLQVCVQT VRTQVYIAVN DKALYEQVVM DYLDNLKPRV EAPKQEEPPN TEDSKTEEKS

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1265       1275       1285       1295       1305       1315
                EMCR        DELGVRVLDQ SDNNCWISTT LIQLQLTKLL DDSIEMQLFK VGKVDSIVQK CYELSHLISG
                229E        ELNGLKILKQ LDNNCWVNSV MLQIQLTGIL DGDYAMQFFK MGRVAKMIER CYTAEQCIRG
                PEDV        SYGGLKVLRQ SHNNCWVTST LVQLQLLGIV D-DPAMELFS AGRVGPMVRK CYESQKAILG
                TGEV        NLNGKIILKQ GDNNCWINAC CYQLQAFDFF N-NEAWEKFK KGDVMDFVNL CYAATTLARG
                OV43        DSHSMAVVDG KQIDDHRITS ITSDKFDFII G-HGMSFSMT TFEIAQLYGS CITP-NVCFV
                BoCoV       DSHSMAVVDG KQIDDHRITS ITSDKFDFII G-HGTSFSMT TFEIAQLYGS CITP-NVCFV
                MHV         DCHSMAVVDG KQIDGKVVTK FNGDKYDFMV G-HGMAFSMS AFEIAQLYGS CITP-NVCFV
                AIBV        QIKQEPIQVV KPQREKKAKK FKVKPATCEK P-------K- FLEYKTCVGD LTVVIAKALD
                SARS CoV    VVQKPVDVKP KIKACIDEVT TTLEETKFLT N-KLLLFADI NGKLYHDSQN MLRGEDMSFL

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  1325       1335       1345       1355       1365       1375
```

```
EMCR       SLGDSGKLLS ELLKDKYTCS ITFEMSCDCG KKFDEQVGCL FWIMPYTKLF QKGECCICHK
229E       AMGDVGLCMY RLLKDLHTGF MVMDYKCSCT SGRLEESGAV LFCTPTKKAF PYGTCLNCNA
PEDV       SLGDVSACLE SLTKDLHTLK ITCSVVCGCG TGERIYEGCA FRMTPTLEPF PYGACAQCAQ
TGEV       HSGDAEYLLE LMLNDYSTAK IVLAAKCGCG EKEIVLERAV FKLTPLKESF NYGVCGDCMQ
OV43       KGDIIKVSKL VKAEVVVNPA NGHMAHGGGV AKAIAVAAGQ QFVKETTDMV KSKGVCATGD
BoCoV      KGDIIKVSKR VKAEVVVNPA NGHMAHGGGV AKAIAVAAGQ QFVKETTDMV KSKGVCATGD
MHV        KGDVIKVLRR VGAEVIVNPA NGRMAHGAGV AGAIAKAAGK SFIKETADMV KNQGVCQVGE
AIBV       EFKEFCIVNA ANEHMTHGSG VAKAIADFCG LDFVEYCEDY VKKHGPQQRL VTPSFVKGIQ
SARS CoV   EKDAPYMVGD VITSGDITCV VIPSKKAGGT TEMLSRALKK VPVDEYITTY PGQGCAGYTL

                 1385       1395       1405       1415       1425       1435
EMCR       MQTYKLVSMK GTGVFVQD-- PAPIDIDAFP VRPICSSVYL GVKGSGHYQT NLYSFDKAID
229E       PRMCTIRQLQ GTIIFVQQK- PEPVNPVSFV VKPVCSSIFR GAVSCGHYQT NIYSQNLCVD
PEDV       VLMHTFKSIV GTGIFCRD-- TTALSLDSLV VKPLCAAAFI GK-DSGHYVT NFYDAAMAID
TGEV       VNTCRFLSVE GSGVFVHDIL SKQTPEAMFV VKPVMHAVYT GTTQNGHYMV DDIEHGYCVD
OV43       CYVSTGGKLC KTVLNVVGPD ARTQGKQSYV LLERVYKHLN NYDCVVTTLI SAGIFSVPSD
BoCoV      CYVSTGGKLC KTVLNVVGPD ARTQGKQSYA LLERVYKHLN KYDCVVTTLI SAGIFSVPSD
MHV        CYESTGGNLC KTVLNIVGPD ARGHGKQCYS FLERAYQHIN KCDDVVTTLI SAGIFSVPTD
AIBV       CVNNVVGPRH GDNNLHEKLV AAYKNVLVDG VVNYVVPVLS LGIFGVDFKM SIDAMREAFE
SARS CoV   EEAKTALKKC KSAFYVLPSE APNAKEEILG TVSWNLREML AHAEETRKLM PICMDVRAIM

                 1445       1455       1465       1475       1485       1495
EMCR       GFGVFDIK-- ---------- ------NSSV NTVCFVDVDF HS-VEIEAGE VK--------
229E       GFGVNKIQP- ---------- ----WTNDAL NTICIKDADY NAKVEISVTP IKNTVDTTPK
PEDV       GYGRHQIK-- ---------- ------YDTL NTICVKDVNW TAPLVPAVDS VVEP------
TGEV       GMGIKPLKKR CYTSTLFINA NVMTRAEKPK QEFKVEKVEQ QPIVEENKSS IEKEEIQSPK
OV43       VSLTYLLGTA KKQVVLVSNN QEDFDLISKC QITAVEG-TK KLAARLSFNV GRSIVYETDA
BoCoV      VSLTYLLGTA KKQVVLVSNN QEDFDLISKC QITAVEG-TK KLAERLSFNV GRSIVYETDA
MHV        VSLTYLIGVV TKNVILVSNN KDDFDVIEKC QVTSIAG-TK ALSLQLAKNL CRDVKFETNA
AIBV       GCTIRVLLFS ---------- ------LSQE HIDYFDVTCK QKTIYLTEDG VKYR------
SARS CoV   ATIQRKYKGI KIQEGIVDYG VRFFFYTSKE PVASIITKLN SLNEPLVTMP IGYVTHGFNL

                 1505       1515       1525       1535       1545       1555
EMCR       ---------- PFAVYKNVKF YLGDISHLVN CVSFDFVVNA ANENLMHGGG VARAIDILTE
229E       EEFVVKEKLN AFLVHDNVAF YQGDVDTVVN GVDFDFIVNA ANENLAHGGG LAKALDVYTK
PEDV       -------VVK PFYSYKNVDF YQGDFSDLVK -LPCDFVVNA ANEKLSHGGG IAKAIDVYTK
TGEV       ND----DLIL PFYKAGKLSF YQGALDVLIN FLEPDVIVNA ANGDLKHMGG VARAIDVFTG
OV43       NKLILIN--- DVAFVSTFNV LQDVLSLRHD IALDDDARTF VQSNVDVVPE GWRVVNKFYQ
BoCoV      NKLILSN--- DVAFVSTFNV LQDVLSLRHD IALDDDARTF VQSNVDVVPE GWRVVNKFYQ
MHV        CDSLFS---- DSCFVSSYDV LQEVELLRHD IQLDDDARVF VQAHMDNLPA DWRLVNKFDS
AIBV       ---------- ---------- ---SIVLKPG DSLGQFGQVY AKNKIVFTAD DVEDKEILYV
SARS CoV   EEAARCMR-- SLKAPAVVSV SSPDAVTTYN GYLTSSSKTS EEHFVETVSL AGSYRDWSYS

                 1565       1575       1585       1595       1605       1615
EMCR       GQLQSLSKDY ISSNGPLKVG AGVMLE--CE KFN--VFNVV GPRTG----- KHEHSLLVEA
229E       GKLQRLSKEH IGLAGKVKVG TGVMVE--CD SLR--IFNVV GPRKG----- KHERDLLIKA
PEDV       GMLQKCSNDY IKAHGPIKVG RGVMLE--AL GLK--VFNVV GPRKG----- KHAPELLVKA
TGEV       GKLTERSKDY LKKNKSIAPG NAVFFENVIE HLS--VLNAV GPRNGD---- SRVEAKLCNV
OV43       INGVRT-VKY FECTGGIDIC SQDKVFGYVQ QGIFNKATVA QIKALF---- LDKVDILLTV
BoCoV      INGVRP-VKY FECPGGIDIC SQDKVFGYVQ QGSFNKATVA QIKALF---- LDKVDILLTV
MHV        VDGVRT-VKY FECPGEIFVS SQGKKFGYVQ NGSFKVASVS QIRALL---- ANKVDVLCTV
AIBV       PTTDKSILEY YGLDAQKYVI YLQTLAQKWN VQYRDNFLIL EWRDGN--CW ISSAIVLLQA
SARS CoV   GQRTELGVEF LKRGDKIVYH TLESPVEFHL DG--EVLSLD KLKSLLSLRE VKTIKVFTTV

                 1625       1635       1645       1655       1665       1675
EMCR       YNSILFENGI PLMPLLSCGI FGVRIENSLK ALFSCDINKP LQVFVYSSNE EQAVLKFLDG
229E       YNTINNEQGT PLTPILSCGI FGIKLETSLE VLLDVCNTKE VKVFVYTDTE VCKVKDFVSG
PEDV       YKSVFANSGV ALTPLISVGI FSVPLEESLS AFLACVGDRH CKCFCYGDKE REAIIKYMDG
TGEV       YKAIAKCEGK ILTPLISVGI FNVRLETSLQ CLLKTVNDRG LNVFVYTDQE RQTIENFFS-
OV43       DGVNFTNRFV PVGESFGKSL GNVFCDGVNV TKHKCDINYK GKVFFQFDNL SSEDLKAVRS
BoCoV      DGVNFTNRFV PVGESFGKSL GNVFCDGVNV TKHKCDINYK GKVFFQFDNL SSEDLKAVRS
MHV        DGVNFRSCCV AEGEVFGKTL GSVFCDGINV TKVRCSAIHK GKVFFQYSGL SAADLVAVTD
AIBV       AKIRFKGFLT EAWAKLLGGD PTDFVAWCYA SCTAKVGDFS DANWLLANLA EHFDADYTNA
SARS CoV   DNTNLHTQLV DMSMTYGQQF GPTYLDGADV TKIKPHVNHE GKTFFVLPSD DTLRSEAFEY

                 1685       1695       1705       1715       1725       1735
EMCR       LDLTPVID-- ------DVDV V---------- --KPFRVEGN FSFFDCG--- VNALDGD-IY
229E       LVNVQKVE-- ------QPKI EPKPVSVIKV APKPYRVDGK FSYFTED--- LLCVADDKPI
PEDV       LVDAIFKEAL VDTTPVQEDV QQVSQKPVLP NFEPFRIEGA HAFYECNPEG LMSLGAD-KL
TGEV       ---------- ---------- ---------- ---------- ---------- ----------
OV43       SFNFDQKELL AYYNMLVN-- CFKWQVVVNG KYFTFKQANN NCFVNVSCLM LQSLHLTFKI
BoCoV      SFNFDQKELL AYYNMLVN-- CSKWQVVFNG KYFTFKQANN NCFVNVSCLM LQSLNLKFKI
MHV        AFGFDEPQLL KYYNMLG--- MCKWPVVVCG NYFAFKQSNN NCYINVACLM LQHLSLKFHK
AIBV       FLKKRVSCN- ---------- ---------- ---------- ---------- ----------
```

```
                SARS CoV    YHTLDESFLG RYMSALNH-- TKKWKFPQVG GLTSIKWADN NCYLSSVLLA LQQLEVKFNA

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1745       1755       1765       1775       1785       1795
                EMCR        LLFTNSILML DKQGQL---- LDTKLNGILQ QAVLDYLATV KTVPAGNLVK LVVE-SCTIY
                229E        VLFTDSMLTL DDRGLA---- LDNALSGVLS AAIKDCVDIN KAIPSGNLIK FDIG-SVVVY
                PEDV        VLFTNSNLDF CSVGKC---- LNDVTSGALL EAINVFKKSN KTVPAGNCVT LDCANMISIT
                TGEV        ---------- ---------- ---------- ---------- ---------- ----------
                OV43        VQWQEAWLEF RSGRPARFVA LVLAKGGFKF GDPADSRDFL RVVFSQVDLT GAICDFEIAC
                BoCoV       VQWQEAWLEF RSGRPARFVS LVLAKGGFKF GDPADSRDFL RVVFSQVDLT GAICDFEIAC
                MHV         WQWQEAWNEF RSGKPLRFVS LVLAKGSFKF NEPSDSTDFM RVVLREADLS GATCDFEFVC
                AIBV        ---------- ---------- ---------- ---------- ---------- ----------
                SARS CoV    PALQEAYYRA RAGDAANFCA LILAYSNKTV GELGDVRETM THLLQHANLE SAKRVLNVVC

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1805       1815       1825       1835       1845       1855
                EMCR        M-CVVPSIND LSFDKNLGRC VRKLNRLKTC VIANVPAIDV LKKLLSSLTL TVKFVVESNV
                229E        M-CVVPSEKD KHLDNNVQRC TRKLNRLMCD IVCTIPADYI LPLVLSSLTC NVSFVGELKA
                PEDV        M-VVLPFDGD ANYDKNYARA VVKVSKLKGK LVLAVDDATL YSKLS--HLS VLGFVSTPDD
                TGEV        --CSIP---- ---------- ---------- ---------- ---------- ----------
                OV43        K-CGVKQEQR TGLDAVMHFG TLSREDLEIG YTVDCSCG-- -KKLIHCVRF DVPFLICSNT
                BoCoV       K-CGVKQEQR TGVDAVMHFG TLSREDLEIG YTVDCSCG-- -KKLIHCVRF DVPFLICSNT
                MHV         K-CGVKQEQR KGVDAVMHFG TLDKGDLAKG YTIACTCG-- -NKLVHCTQL NVPFLICSNK
                AIBV        --CGIKSYEL RGLEACIQP- ---------- ---------- ---------V RATN------
                SARS CoV    KHCGQKTTTL TGVEAVMYMG TLSYDNLKTG VSIPCVCGR- -DATQYLVQQ ESSFVMMSAP

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1865       1875       1885       1895       1905       1915
                EMCR        MDVNDCFKND NVVLKITEDG INVKDVVVES SKSLGKQLG- VVSDGVDSFE GVLP--INTD
                229E        AEA------K VITIKVTEDG VNVHDVTVTT DKSFEQQVG- VIADKDKDLS GAVPSDLNTS
                PEDV        VER--FYANK SVVIKVTEDT RSVKAVKVES TATYGQQIG- PCLVNDTVVT DNKP--VVAD
                TGEV        --VN------ -----VTEDN VNHERVSVSF DKTYGEQLKG TVVIKDKDVT NQLPSAFDVG
                OV43        PASVKLPKG- VGSANIFIG- DKVGHYVHVK CEQSYQLYDA SNVKKVTDVT GKLSDCLYLK
                BoCoV       PASVKLPKG- VGSANIFKG- DKVGHYVHVK CEQSYQLYDA SNVKKVTDVT GKLSDCLYLK
                MHV         PEGKKLPDD- VVAANIFTG- GSLGHYTHVK CKPKYQLYDA CNVSKVSEAK GNFTDCLYLK
                AIBV        -----LLHFK TQYSNCPTCG ANNTDEVIEA SLPYLLLFAT DGPATVDCDE DAVG------
                SARS CoV    PAEYKLQQGT FLCANEYTGN YQCGHYTHIT AKETLYRIDG AHLTKMSEYK GPVTDVFY-K

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1925       1935       1945       1955       1965       1975
                EMCR        TVLSVAPEVD WVAFYGFEKA ALFASLDVKP YGYPNDFVGG FRVLGTTDNN CWVNATCIIL
                229E        ELLTKAIDVD WVEFYGFKDA VTFATVDHSA FAYESAVVNG IRVLKTSDNN CWVNAVCIAL
                PEDV        VVAKVVPNAN WDSHYGFDKA GEFHMLDHTG FTFPSEVVNG RRVIKTTDNN CWVNVTCLQL
                TGEV        QKVIKAIDID WQAHYGFRDA AAFSASSHDA YKFEVVTHSN FIVHKQTDNN CWINAICLAL
                OV43        NLKQTFKSVL TTYYLDDVKK IEYKPDLSQY YCDGGKYYTQ RIIKAQFKTF EKVDGVYTNF
                BoCoV       NLKQTFKSVL TTYYLDDVKK IEYKPDLSQY YCDGGKYYTQ RIIKAQFKTF EKVDGVYTNF
                MHV         NLKQTFSSKL TTFYLDDVKC VEYNPDLSQY YCESGKYYTK PIIKAQFRTF EKVEGVYTNF
                AIBV        TVVFVGSTNS GHCYTQAAGQ AFDNLAKDRK FGKKSPYITA MYTRFAFKNE TSLPVAKQSK
                SARS CoV    ETSYTTTIKP VSYKLDGVTY TEIEPKLDGY YKKDNAYYTE QPIDLVP-TQ PLPNASFDNF

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              1985       1995       2005       2015       2025       2035
                EMCR        QYLKPTFKSK GLNVLWNKFV TGDVGPFVSF IYFITMSSKG QKGDAEEALS KLSEYLISDS
                229E        QYSKPHFISQ GLDAAWNKFV LGDVEIFVAF VYYVARLMKG DKGDAEDTLT KLSKYLANEA
                PEDV        QFARFRFKSA GLQAMWESYC TGDVAMFVHW LYWLTGVDKG QPSDSENALN MLSKYIVPAG
                TGEV        QRLKPQWKFP GVRGLWNEFL ERKTQGFVHM LYHISGVKKG EPGDAELMLH KLGDLMDNDC
                OV43        KLIG--HTVC DSLNAKLGFD SSKEFVEYKI TEWPTATGDV VLATDDLYVK RYERGCITFG
                BoCoV       KLIG--HTVC DILNAKLGFD SSKEFVEYKV TEWPTATGDV VLATDDLYVK RYERGCITFG
                MHV         KLVG--HSIA EKFNAKLGFD CNSPFTEYKI TEWPTATGDV VLATDDLYVK RYERGCITFG
                AIBV        GKSKS-VKED VSNLATSSKA SFDNLTDFEQ WYDSNIYESL KVQESPDNFD KYVSFTTKED
                SARS CoV    KLTCSNTKFA DDLNQMTGFT KP-ASRELSV TFFPDLNGDV VAIDYRHYSA SFKKGAKLLH

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              2045       2055       2065       2075       2085       2095
                EMCR        IVTLEQYSTC DIC------- -----KSTVV EVKSAVVCAS VLKDGCDVG- ----------
                229E        QVQLEHYSSC VECDAKF--- -----KNSVA SINSAIVCAS VKRDGVQVG- ----------
                PEDV        SVTIERVTHD GCC------- -------CSKR VVTAPVVNAS VLKLGVEDG- ----------
                TGEV        EIIVTHTTAC DKC------- -------AKVE KFVGPVVAAP LAIHGTDE-- ----------
                OV43        KPVIWLSHEK ASLNSLT--- -----YFNRP SLVDDNKFDV LKVDDVD--- ----------
                BoCoV       KPVIWLSHEQ ASLNSLT--- -----YFNRP LLVDENKFDV LKVDDVD--- ----------
                MHV         KPVIWLGHEE ASLKSLT--- -----YFNRP SVVCENKFNV LPVDVSEPTD KGPVPAAVLV
                AIBV        SKLPLTLKVR GIKS------ ---------VV DFRSKDGFIY KLTPDTD--- ----------
                SARS CoV    KPIVWHINQA TTKTTFKPNT WCLRCLWSTK PVDTSNSFEV LAVEDTQGMD N---------

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              2105       2115       2125       2135       2145       2155
                EMCR        ----FCPHRH KLRSRVK--- ---------- ---------- -FVNGRVVIT NVGEPIISQP
                229E        ----YCVHGI KYYSRVR--- ---------- ---------- -SVRGRAIIV SVEQLEPCAQ
                PEDV        ----LCPHGL NYIGKVV--- ---------- ---------- -VVKGTTIVV NVGKPVVAPS
                TGEV        ----TCVHGV SVNVKVT--- ---------- ---------- -QIKGTVAIT SLIGPIIG--
```

```
                OV43        ----DGGDSS ESGAKE---- ---------- ---------- TKEINIIKLS GVKKPFKVED
                BoCoV       ----DGGDIS ESDAKE---- ---------- ---------- PKEINIIKLS GVKKPFKVED
                MHV         TGALSGAATA PGTAKEQKVC ASDSVVDQVV SGFLSDLSGA TVDVKEVKLN GVKKPIKVED
                AIBV        ---------- ---------- ---------- ---------- --ENSKAPVY YPVLDAISLK
                SARS CoV    ----LACESQ QPTSEEVVEN ---------- ---------- -PTIQKEVIE CDVKTTEVVG

                                 2165       2175       2185       2195       2205       2215
                EMCR        SKLLNGIAYT TFS--GSFDN GHYVVYDAAN NAVYDGARLF ASDLSTLAVT AIVVVGGCVT
                229E        SRLLSGVAYT AFS--GPVDK GHYTVYDTAK KSMYDGDRFV KHDLSLLSVT SVVMVGGYVA
                PEDV        HLFLKGVSYT TFLDNGNGVV GHYTVFDHGT GMVHDGDAFV PGDLNVSPVT NVVVSEQTAV
                TGEV        -EVLEATGYI CYS--GSNRN GHYTYYDNRN GLVVDAEKAY HFNRDLLQVT TAIASNFVVK
                OV43        SVIVNDDTSE TKYVKSLSIV DVYDMWLTGC KYVVRTANAL SRAVNVPTIR KFIKFGMTLV
                BoCoV       SVIVNDDTSE IKYVKSLSIV DVYDMWLTGC RCVVRTANAL SRAVNVPTIR KFIKFGMTLV
                MHV         SVVVNDPTSE TKVVKSLSIV DVYDMFLTGC RYVVWMANEL SRLVNSPTVR EYVKWGMTKI
                AIBV        AIWVEGNANF VVG-----HP NYYSKSLHIP TFWENAENFV KMGDKIGGVT MGLWRAEHLN
                SARS CoV    NVILKPSDEG VKVTQELGHE DLMAAYVENT SITIKKPNEL SLALGLKTIA THGIAAINSV

                                 2225       2235       2245       2255       2265       2275
                EMCR        S--------- ------NVPP IVSEKISVMD KLDTG---AQ KFFQFGDFVM NNIVLFLTWL
                229E        ---------- ---------PV NTVKPKPVIN QLDEK---AQ KFFDFGDFLI HNFVIFFTWL
                PEDV        V--------- ------IKDP VKKAELDATK LLDTMNYASE RFFSFGDFMS RNLITVFLYI
                TGEV        KPQAEERPKN CAFNKVAASP KIVQEQKLLA IESGANYALT EFGRYADMFF MAGDKILRLL
                OV43        SIP------- ------IDLL NLREIKPAVN VVKAVRNKIS VCFNFIKWLF VLLFGWIKIS
                BoCoV       SIP------- ------IDLL NLREIKPVFN VVKAVRNKIS ACFNFIKWLF VLLFGWIKIS
                MHV         VIP------- ------AKLV LLRDEKQEFV APKVVKAKVI ACYSAVKWFF LYCFSWIKFN
                AIBV        KPN------- ------LERI FNIAKKAIVG SSVVTTQCGK LIGKAATFIA DKVGGGVVRN
                SARS CoV    PWS------- ------KILA YVKPFLG--- --QAAITTSN CAKRLAQRVF NNYMPYVFTL

                                 2285       2295       2305       2315       2325       2335
                EMCR        LSMFSLLRTS IMKHDIKVIA KAPKRTGVIL TRSFKYNIRS ALFVVKQKWC VIVTLFKFLL
                229E        LSMFTLCKTA VTTGDVKIMA KAPQRTGVVL KRSLKYNLKA SAAVLKSKWW LLAKFTKLLL
                PEDV        LSILGLCFRA FRKRDVKVLA GVPQRTGIIL RKSMRYNAKA LGVFFKLKLY WFKVLGKFSL
                TGEV        LEVFKYLLVL FMCLRSTKMP KVKVKP-PLA FKDFGAKVRT LNYMRQLNKP SVWRYAKLVL
                OV43        ADNKVIYTTE IASKLTCKLV ALAFKNAFLT FKWSMVARGA CIIATIFLLW FNFIYANVIF
                BoCoV       ADNKVIYTTE VASKLTCKLV ALAFKNAFLT FKWSVVARGA CIIATIFLLW FNFIYANVIF
                MHV         TDNKVIYTTE VASKLTFNLC CLAFKNALQT FNWNVVSRGF FLVATVFLLW FNFLYANVIL
                AIBV        ITDSIKGLCG ITRGHFERKM SPQFLKTLMF FLFYFLKASV KSVVASYKTV LCKVVLATLL
                SARS CoV    LFQLCTFTKS TNSRIRASLP TTIAKNSVKS VAKLCLDAGI NYVKSPKFSK LFTIAMWLLL

                                 2345       2355       2365       2375       2385       2395
                EMCR        LLYAIYALVF MIVQFSPFNS LLCGDIVSGY EKSTFN---- --KDIYCGNS MVCKMCLFSY
                229E        LIYTLYSVVL LCVRFGPFN- -FCSETVNGY AKSNFV---- --KDDYCDGS LGCKMCLFGY
                PEDV        GIYALYALLF MTIRFTPIGS PVCDDVVAGY ANSSFD---- --KNEYCN-S VICKVCLYGY
                TGEV        LLIAIYNFFY LFVSIPVVHK LTCNGAVQAY KNSSFI---- --KSAVCGNS ILCKACLASY
                OV43        SDFYLPKIGF LPTFVGKIAQ WIKNTFSLVT ICDLYSMQDV GFKNQYCNGS IACQFCLAGF
                BoCoV       SDFYLPKIGF LPTFVGKIVQ WIKNTFSLVT ICDLYSIQDV GFKNQYCNGS IACQFCLAGF
                MHV         SDFYLPNIGF FPTFVGQIVA WVKTTFGIFT LCDLYQVSDV GYRSSFCNGS MVCELCFSGF
                AIBV        IVWFVYTSNP VMFTGIRVLD FLFEGSLCGP YKDYGK--DS FDVLRYCADD FICRVCLHDK
                SARS CoV    LSICLGSLIC VTAAFGVLLS NFGAPSYCNG VRELYLNSSN VTTMDFCEGS FPCSICLSGL

                                 2405       2415       2425       2435       2445       2455
                EMCR        QEFNDLDHTS LVWKHIR--- ----D--P-- -ILISLQPFV ILVILLIFGN MYLRFGLLYF
                229E        QELSQFSHLD VVWKHIT--- ----D--P-- -LFSNMQPFI VMVLLLIFGD NYLRCFLLYF
                PEDV        QELSDFSHTQ VVWQHLR--- ----D--P-- -LIGNVMPFF YLAFLAIFGG VYVKAITLYF
                TGEV        DELADFQHLQ VTWDFKS--- ----D--P-- -LWNRLVQLS YFAFLAVFGN NYVRCFLMYF
                OV43        DMLDNYKAID VVQYEADRRA FVDYTGVLKI VIELIVSYAL YTAWFYPLFA LISIQILTTW
                BoCoV       DMLDNYKAID VVQYEADRRA FVDYTGVLKI VIELIVSYAL YTAWFYPLFA LISIQILTTW
                MHV         DMLDNYDAIN VVQHVVDRRV SFDYISLFKL VVELVIGYSL YTVCFYPLFG LIGMQLLTTW
                AIBV        DSLHLYKHAY SVEQVYKDAA SG-------- --FIFNWNWL YLVFLILFVK PVAGFVIICY
                SARS CoV    DSLDSYPALE TIQVTIS--S YKLDLTILGL AAEWVLAYML FTKFFYLLGL SAIMQVFFGY

                                 2465       2475       2485       2495       2505       2515
                EMCR        VAQFISTFG- -SFLGFHQKQ WFLHFVPFDV LCNEFLATFI VCKIVLFVRH IIVGCNNADC
                229E        VAQMISTVG- -VFLGYKETN WFLHFIPFDV ICDELLVTVI VIKVISFVRH VLFGCENPDC
                PEDV        IFQYLNSLG- -VFLGLQQSI WFLQLVPFDV FGDEIVVFFI VTRVLMFIKH VCLGCDKASC
                TGEV        VSQYLNLWL- -SYFGYVEYS WFLHVVNFES ISAEFVIVVI VVKAVLALKH IVFACSNPSC
                OV43        LPELFMLST- -LHWSFRLLV ALANMLPAHV FMRFYIIIAS FIKLFSLFRH VAYGCSKSGC
                BoCoV       LPELLMLST- -LHWSVRLLV SLANMLPAHV FMRFYIIIAS FIKLFSLFRH VAYGCSKSGC
                MHV         LPEFFMLET- -MHWSARFFV FVANMLPAFT LLRFYIVVTA MYKIFCLCRH VMYGCSRPGC
                AIBV        CVKYLVLNST VLQTGVCFLD WFVQTVFSHF NFMGAGFYFW LFYKIYIQVH HILYCKDVTC
                SARS CoV    FASHFISN-- --SWLMWFII SIVQMAPVSA MVRMYIFFAS FYYIWKSYVH IMDGCTSSTC

                                 2525       2535       2545       2555       2565       2575
```

```
                EMCR        VACSKSARLK RVPLQTIING MHKSFYVNAN GGTCFCNKHN FFCVNCDSFG PGNTFINGDI
                229E        IACSKSARLK RFPVNTIVNG VQRSFYVNAN GGSKFCKKHR FFCVDCDSYG YGSTFITPEV
                PEDV        VACSKSARLK RVPVQTIFQG TSKSFYVHAN GGSKFCKKHN FFCLNCDSYG PGCTFINDVI
                TGEV        KTCSRTARQT RIPIQVVVNG SMKTVYVHAN GTGKFCKKHN FYCKNCDSYG FENTFICDEI
          5     OV43        LFCYKRNRSL RVKCSTIVGG MIRYYDVMAN GGTGFCSKHQ WNCIDCDSYK PGNTFITVEA
                BoCoV       LFCYKRNRSL RVKCSTIVGG MIRYYDVMAN GGTGFCSKHQ WNCIDCDSYK PGNTFITVEA
                MHV         LFCYKRNRSV RVKCSTVVGG TLRYYDVMAN GGTGFCAKHQ WNCLNCSAFG PGNTFITHEA
                AIBV        EVCKRVARSN RQEVSVVVGG RKQIVHVYTN SGYNFCKRHN WYCRNCDDYG HQNTFMSPEV
                SARS CoV    MMCYKRNRAT RVECTTIVNG MKRSFYVYAN GGRGFCKTHN WNCLNCDTFC TGSTFISDEV
          10

                                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                                  2585        2595        2605        2615        2625        2635
                EMCR        ARELGNVVKT AVQPTAPAYV IIDKVDFVNG FYRLYSGDTF WRYDFDITES KYSCKE----
                229E        SRELGNITKT NVQPTGPAYV MIDKVEFENG FYRLYSCETF WRYNFDITES KYSCKE----
          15    PEDV        ATEVGNVVKL NVQPTGPATI LIDKVEFSNG FYYLYSGDTF WKYNFDITDS KYTCKE----
                TGEV        VRDLSNSVKQ TVYATDRSHQ EVTKVECSDG FYRFYVGDEF TSYDYDVKHK KYSSQE----
                OV43        ALDLSKELKR PIQPTDVAYH TVTDVKQVGC SMRLFYDRDG QRTYDDVNAS LFVDYSNLLH
                BoCoV       ALDLSKELKR PIQPTDVAYH TVTDVKQVGC YMRLFYDRDG QRTYDDVNAS LFVDYSNLLH
                MHV         AADLSKELKR PVNPTDSAYY LVTEVKQVGC SMRLFYERDG QRVYDDVSAS LFVDMNGLLH
          20    AIBV        AGELSEKLKR HVKPTAYAYH VVDEACLVDD FVNLKYKAAT PGKDSASSAV KCFSVTDFLK
                SARS CoV    ARDLSLQFKR PINPTDQSSY IVDSVAVKNG ALHLYFDKAG QKTYERHPLS HFVNLDNLRA

                                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
          25                    2645        2655        2665        2675        2685        2695
                EMCR        -VLKNCNVLE NFIVYNN--- -----SGSNI TQIKNACVYF SQLLCEPIKL VNSELLSTLS
                229E        -VFKNCNVLD DFIVFNN--- -----NGTNV TQVKNASVYF SQLLCRPIKL VDSELLSTLS
                PEDV        -ALKNCSIIT DFIVFNN--- -----NGSNV NQVKNACVYF SQMLCKPVKL VDSALLASLS
                TGEV        -VLKSMLLLD DFIVYSP--- -----SGSAL ANVRNACVYF SQLIGKPIKI VNSDLLEDLS
                OV43        SKVKSVPNMH VVVVEN---- ------DADK ANFLNAAVFY AQSLFRPILM VDKNLITTAN
          30    BoCoV       SKVKSVPNMH VVVVEN---- ------DADK ANFLNAAVFY AQSLFRPILM VDKILITTAN
                MHV         SKVKGVPETH VVVVEN---- ------EADK AGFLNAAVFY AQSLYRPMLL VEKKLITTAN
                AIBV        KAVFLKEALK CEQISNDGFI VCNTQSAHAL EEAKNAAIYY AQYLCKPILI LDQALYEQLV
                SARS CoV    NNTKGSLPIN VIVFDGK--- -----SKCDE SASKSASVYY SQLMCQPILL LDQVLVSDVG
          35

                                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                                  2705        2715        2725        2735        2745        2755
                EMCR        --VDFNGVLH KAYVDVLCNS FFKELTANMS MAECKATLGL T---------- ----------
                229E        --VDFNGVLH KAYIDVLRNS FGKDLNANMS LAECKRALGL S---------- ----------
                PEDV        --VDFGASLH SAFVSVLSNS FGKDLSSCND MQDCKSTLGF DD---------- ----------
          40    TGEV        --VDFKGALF NAKKNVIKNS FNVDVSECKN LDECYRACNL N---------- ----------
                OV43        TGTSVTETMF DVYVDTFLSM FDVDKKSLNA LIATAHSSIK QGTQIYKVLD TFLSCARKSC
                BoCoV       TGTSVTETMF DVYVDTFLSM FDVDKKSLNA LIATAHSSIK QGTQICKVLD TFLSCARKSC
                MHV         TGLSVSQTMF DLYVDSLLGV LDVDRKSLTS FVNAAHNSLK EGVQLEQVMD TFIGCARRKC
                AIBV        V-EPVSKSVI DKVCSILSSI ISVDTAALNY KAGTLRDALL S---------- ----------
          45    SARS CoV    DSTEVSVKMF DAYVDTFSAT FSVPMEKLKA LVATAHSELA KGVALDGVLS TFVSAARQG-

                                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                                  2765        2775        2785        2795        2805        2815
                EMCR        -----VSDDD FVSAVANAHR YDVLLSDLSF NNFFISYAKP EDK-LSVYDI ACCMRAGSKV
          50    229E        -----ISDHE FTSAISNAHR CDVLLSDLSF NNFVSSYAKP EEK-LSAYDL ACCMRAGAKV
                PEDV        -----VPLDT FNAAVAEAHR YDVLLTDMSF NNFTTSYAKP EEK-FPVHDI ATCMRVGAKI
                TGEV        -----VSFST FEMAVNNAHR FGILITDRSF NNFWPSKVKP GSSGVSAMDI GKCMTSDAKI
                OV43        SIDSDVDTKC LADSVMSAVS AGLELTDESC NNLVPTYLKS DN--IVAADL GVLIQNSAKH
          55    BoCoV       SIDSDVDTKC LADSVMSAVS AGLELTDESC NNLVPTYLKG DN--IVAADL GVLIQNSAKH
                MHV         AIDSDVETKS ITKSIMSAVN AGVDFTDESC NNLVPTYVKS DT--IVAADL GVLIQNNAKH
                AIBV        ----ITKDEE AVDMAIFCHN HDVDYTGDGF TNVIPSYGID TG-KLTPRDR GFLINADASI
                SARS CoV    VVDTDVDTKD VIECLKLSHH SDLEVTGDSC NNFMLTYNKV EN--MTPRDL GACIDCNARH

          60                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                                  2825        2835        2845        2855        2865        2875
                EMCR        VNHNVLIKES IPIVWGVKDF NTLSQEGKKY LVKTTKAKGL TFLLTFNDNQ AITQVP----
                229E        VNANVLTKDQ TPIVWHAKDF NSLSAEGRKY IVKTSKAKGL TFLLTINENQ AVTQIP----
                PEDV        VNHNVLVKDS IPVVWLVRDF IALSEETRKY IIRTTKVKGI TFMLTFNDCR MHTTIP----
                TGEV        VNAKVLTQRG KSVVWLSQDF AALSSTAQKV LVKTFVEEGV NFSLTFNAVG SDDDLPYERF
          65    OV43        VQGNVAKIAG VSCIWSVDAF NQFSSDFQHK LKKACCKTGL KLKLTYNKQM ANVSVLT---
                BoCoV       VQGNVAKIAG VSCIWSVDAF NQLSSDFQHK LKKACCKTGL KLELTYNKQM ANVSVLT---
                MHV         VQANVAKAAN VACIWSVDAF NQLSADLQHR LRKACSKTGL KIKLTYNKQE ANVPILT---
                AIBV        ANLRVKN--A PPVVWKFSEL IKLSDSCLKY LISATVKSGV RFFITKSGAK QVIACHT---
                SARS CoV    INAQVAKSHN VSLIWNVKDY MSLSEQLRKQ IRSAAKKNNI PFRLTCATTR QVVNVIT---
          70

                                ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                                  2885        2895        2905        2915        2925        2935
                EMCR        ATSIVAKQGA G---------- --FKRTYNFL WYVCLFVVAL FIGVSFID-- ---YTTTVTS
                229E        ATSIVAKQGA GD--------- --AGHSLTWL WLLCGLVCLI QFYLCFFMPY --FMYDIVSS
          75    PEDV        TVCIANKKGA GLPS------ --FSKVKKFF WFLCLFIVAA FFALSFLD-- ---FSTQVSS
                TGEV        TESVSPKSGS G---------- --FFDVITQL KQIVILVFVF IFICGLCSVY SVATQSYIES
                OV43        --TPFSLKGG A--V------- --FSYFVYVC FVLSLVCFIG ---LWCLMPT YTVHKSDFQL
                BoCoV       --TPFSLKGG A--V------- --FSYFVYVC FVLSLVCFIG ---LWCLMPT YTVHKSDFQL
                MHV         --TPFSLKGG A--V------- --FSKVLQWL FVVNLICFIV ---LWALMPT YAVHKSDMQL
          80    AIBV        QKLLVEKKAG GIVSGTFKCF KSYFKWLLIF YILFTACCSG YYYMEVSKSF VHPMYDVNST
```

```
                SARS CoV    --TKISLKGG K--I------ --VSTCFKLM LKATLLCVLA ALVCYIVMPV HTLSIHDGYT

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  2945       2955       2965       2975       2985       2995
                EMCR        FHGYDFKYIE NGQLKVFEAP LHCVRNVFDN FNQWHEAKFG VVTTNSDKCP IVVG----VS
                229E        FEGYDFKYIE NGQLKNFEAP LKCVRNVFEN FEDWHYAKFG FTPLNKQSCP IVVG----VS
                PEDV        DSDYDFKYIE SGQLKTFDNP LSCVHNVFIN FDQWHDAKFG FTPVNNPSCP IVVG----VS
                TGEV        AEGYDYMVIK NGIVQPFDDT ISCVHNTYKG FGDWFKAKYG FIPTFGKSCP IVVGT-VFDL
                OV43        PVYASYKVLD NGVIRDVSVE DVCFANKFEQ FDQWYESTFG LSYYSNSMAC PIVVA-VIDQ
                BoCoV       PVYASYKVLD NGVIRDVSVE DVCFANKFEQ FDQWYESTFG LSYYSNSMAC PIVVA-VVDQ
                MHV         PLYASFKVID NGVLRDVTVT DACFANKFIQ FDQWYESTFG LVYYRNSRAC PVVVA-VIDQ
                AIBV        LHVEGFKVID KGVLREIVPE DTCFSNKFVN FDAFWGRPYD NSRNCPIVTA VIDGDGTVAT
                SARS CoV    NEIIGYKAIQ DGVTRDIIST DDCFANKHAG FDAWFSQRGG S--YKNDKSC PVVAA-IITR

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3005       3015       3025       3035       3045       3055
                EMCR        ERINVVPGVP TNVYLVG--- -KTLVFTLQA AFGNTGVCYD FDGVTTS--- --DKCIFNSA
                229E        EIVNTVAGIP SNVYLVG--- -KTLIFTLQA AFGNAGVCYD IFGVTTP--- --EKCIFTSA
                PEDV        DEARTVPGIP AGVYLAG--- -KTLVFAINT IFGTSGLCFD ASGVADK--- --GACIFNSA
                TGEV        ENMRPIPDVP AYVSIVG--- -RSLVFAINA AFGVTNMCYD HTGNAVSKDS YFDTCVFNTA
                OV43        DFGSTVFNVP TKVLRYG--- YHVLHFITHA LSADGVQCYT PHSQISYSNF YASGCVLSSA
                BoCoV       DFGSTVFNVP TKVLRYG--- YHVLHFITHA LSADGVQCYT PHSQISYSNF YASGCVLSSA
                MHV         DIGYTLFNVP TKVLRYG--- FHVLHFITHA FATDSVQCYT PHMQIPYDNF YASGCVLSSL
                AIBV        GVPGFVSWVM DGVMFIHMTQ TERKPWYIPT WFNREIVGYT QDSIITEGSF YTSIALFSAR
                SARS CoV    EIGFIVPGLP GTVLRAIN-- GDFLHFLPRV FSAVGNICYT PSKLIEYSDF ATSACVLAAE

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3065       3075       3085       3095       3105       3115
                EMCR        CTRLEGLGGD -NVYCYNTDL IEGSKPYSIL QPNAYYKYDV K-NYVRFPEI LARGFGLRTI
                229E        CTRLEGLGGN -NVYCYNTAL MEGSLPYSSI QANAYYKYDN G-NFIKLPEV IAQGFGFRTV
                PEDV        CTTLSGLGGT -AVYCYKNGL VEGAKLYSEL APHSYYKMVD G-NAVSLPEI ISRGFGIRTI
                TGEV        CTTLTGLGGT -IVYCAKQGL VEGAKLYSDL MPDYYYEHAS G-NMVKLPAI IR-GLGLRFV
                OV43        CTMFTMADGS PQPYCYTEGL MQNASLYSSL VPHVRYNLAN AKGFIRFPEV LREGL-VRIV
                BoCoV       CTMFAMADGS PQPYCYTDGL MQNASLYSSL VPHVRYNLAN AKGFIRLPEV LREGL-VRIV
                MHV         CTMLAHADGT PHPYCYTEGI MHNASLYDSL APHVRYNLAN SNGYIRFPEV VSEGI-VRIV
                AIBV        CLYLTASNTP QLYCFNGDND APGALPFGSI IPHRVYFQPN GVRLIVPQQI LHTPY---VV
                SARS CoV    CTIFKDAMGK PVPYCYDTNL LEGSISYSEL RPDTRYVLMD G-SIIQFPNT YLEGS-VRVV

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3125       3135       3145       3155       3165       3175
                EMCR        RTLATRYCRV GECRDSHKGV CFGFDKWYVN DGRVD---DG YICGDGLIDL LVNVLSIFSS
                229E        RTIATKYCRV GECVESNAGV CFGFDKWFVN DGRVA---NG YVCGTGLWNL VFNILSMFSS
                PEDV        RTKAMTYCRV GQCVQSAEGV CFGADRFFVY NAESG---SD FVCGTGLFTL LMNVISVFSK
                TGEV        KTQATTYCRV GECIDSKAGF CFGGDNWFVY DNEFG---NG YICGNSVLGF FKNVFKLFNS
                OV43        RTRSMSYCRV GLCEEADEGI CFNFNGSWVL NNDYYRSLPG TFCGRDVFDL IYQLFKGLAQ
                BoCoV       RTRSMSYCRV GLCEEADEGI CFNFNGSWVL NNDYYRSLPG TFCGRDVFDL IYQLFKGLAQ
                MHV         RTRSMTYCRV GLCEDAEEGV CFNFNSSWVL NNPYYRAMPG TFCGRNAFDL IHQVLGGLVR
                AIBV        KFVSDSYCRG SVCEYTRPGY CVSLNPQWVL FNDEYTSKPG VFCGSTVREL MFSMVSTFFT
                SARS CoV    TTFDAEYCRH GTCERSEVGI CLSTSGRWVL NNEHYRALSG VFCGVDAMNL IANIFTPLVQ

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3185       3195       3205       3215       3225       3235
                EMCR        SFSVVAMSGH MLFNFLFAAF ITFLCFLVTK FKRVFGDLSY GVFTVVCATL INNISYVVTQ
                229E        SFSVAAMSGQ ILLNCALGAF AIFCCFLVTK FRRMFGDLSV GVCTVVVAVL LNNVSYIVTQ
                PEDV        TVPVTVLSGQ ILFNCIIAFV AVAVCFLFTK FKRMFGDMSV GVFTVGACTL LNNVSYIVTQ
                TGEV        NMSVVATSGA MLVNIIIACL AIAMCYGVLK FKKIFGDCTF LIVMIIVTLV VNNVSYFVTQ
                OV43        PVDFLALTAS SIAGAILAVI VVLVFYYLIK LKRAFGDYTS VVFVNVIVWC VNFMMLFVFQ
                BoCoV       PVDFLALTAS SIAGAILAVI VVLGFYYLIK LKRAFGDYTS IVFVNVIVWC VNFMMLFVFQ
                MHV         PIDFFALTAS SVAGAILAII VVLAFYYLIK LKRAFGDYTS VVVINVIVWC INFLMLFVFQ
                AIBV        GVN-PNIYMQ LATMFLILVV VVLIFAMVIK FQGVFKAYAT TVFITMLVWV INAFILCVHS
                SARS CoV    PVGALDVSAS VVAGGIIAIL VTCAAYYFMK FRRVFGEYNH VVAANALLFL MSFTILCLVP

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3245       3255       3265       3275       3285       3295
                EMCR        N-LFFMLLYA ILYFVFTRTV R--YAWIWHI AYIVAYFLLI PWWLLTWFSF AAFLELLPNV
                229E        N-LVTMIAYA ILYFFATRSL R--YAWIWCA AYLIAYISFA PWWLCAWYFL AMLTGLLPSL
                PEDV        N-TLGMLGYA TLYFLCTKGV R--YMWIWHL GFLISYILIA PWWVLMVYAF SAIFEFMPNL
                TGEV        N-TFFMIIYA IVYYFITRKL A--YPGILDA GFIIAYINMA PWYVITAYIL VFLYDSLPSL
                OV43        VYPILSCVYA ICYFYATLYF PSEISVIMHL QWLVMYGTIM PLWFCLLYIA VVVSNHAFWV
                BoCoV       VYPTLSCVYA ICYFYATLYF PSEISVIMHL QWLVMYGTIM PLWFCLLYIS VVVSNHAFWV
                MHV         VYPTLSCLYA CFYFYTTLYF PSEISVVMHL QWLVMYGAIM PLWFCIIYVA VVVSNHALWL
                AIBV        YNSVLAVILL VLYCYASLVT SRNTVIIMHC WLVFTFGLIV PTWLACCYLG FIIYMYTPLF
                SARS CoV    AYSFLPGVYS VFYLYLTFYF TNDVSFLAHL QWFAMFSPIV PFWITAIYVF CISLKHCHWF

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  3305       3315       3325       3335       3345       3355
                EMCR        FKLKISTQ-- --LFEGDKFI GTFESAAAGT FVLDMRSYER LINT--ISPE KLKNYAASYN
                229E        LKLKVSTN-- --LFEGDKFV GTFESAAAGT FVIDMRSYEK LANS--ISPE KLKSYAASYN
                PEDV        FKLKVSTQ-- --LFEGDKFV GSFENAAAGT FVLDMHAYER LANS--ISTE KLRQYASTYN
                TGEV        FKLKVSTN-- --LFEGDKFV GNFESAAMGT FVIDMRSYET IVNS--TSIA RIKSYANSFN
```

```
                    OV43        FSYCRKLG-- --TSVR--SD GTFEEMALTT FMITKDSYCK LKNS--LSDV AFNRYLSLYN
                    BoCoV       FSYCRQLG-- --TSVR--SD GTFEEMALTT FMITKDSYCK LKNS--LSDV AFNRYLSLYN
                    MHV         FSYCRKLG-- --TEVR--SD GTFEEMSLTT FMITKESYCK LKNS--VSDV AFNRYLSLYN
                    AIBV        LWCYGTTKNT RKLYDGNEFV GNYDLAAKST FVIRGSEFVK LTNE---IGD KFEAYLSAYA
                    SARS CoV    FNNYLRKR-- --VMFNGVTF STFEEAALCT FLLNKEMYLK LRSETLLPLT QYNRYLALYN

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3365       3375       3385       3395       3405       3415
                    EMCR        KYKYYSGSAS EADYRCACYA HLAKAMLDYA -KDHNDMLYS PPTISYN-ST LQSGLKKMAQ
                    229E        RYKYYSGNAN EADYRCACYA YLAKAMLDFS -RDHNDILYT PPTVSYG-ST LQAGLRKMAQ
                    PEDV        KYKYYSGSAS EADYRLACFA HLAKAMMDYA -SNHNDTLYT PPTVSYN-ST LQAGLRKMAQ
                    TGEV        KYKYYTGSMG EADYRMACYA HLGKALMDYS -VNRTDMLYT PPTVSVN-ST LQSGLRKMAQ
                    OV43        KYRYYSGKMD TAAYREAACS QLAKAMDTFT NNNGSDVLYQ PPTASVSTSF LQSGIVKMVN
                    BoCoV       KYRYYSGKMD TAAYREAACS QLAKAMDTFT NNNGSDVLYQ PPTASVSTSF LQSGIVKMVN
                    MHV         KYRYFSGKMD TAAYREAACS QLAKAMETFN HNNGNDVLYQ PPTASVTTSF LQSGIVKMVF
                    AIBV        RLKYYSGTGS EQDYLQACRA WLAYALDQYR -NSGVEIVYT PPRYSIGVSR LQSGFKKLVS
                    SARS CoV    KYKYFSGALD TTSYREAACC HLAKALNDFS -NSGADVLYQ PPQTSITSAV LQSGFRKMAF

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3425       3435       3445       3455       3465       3475
                    EMCR        PSGCVERCVV RVCYGSTVLN GVWLGDTVTC PRHVIAPSTT VL-IDYDHAY STMRLHNFSV
                    229E        PSGFVEKCVV RVCYGNTVLN GLWLGDIVYC PRHVIASNTT SA-IDYDHEY SIMRLHNFSI
                    PEDV        PSGVVEKCIV RVCYGNMALN GLWLGDIVMC PRHVIASSTT ST-IDYDYAL SVLRLHNFSI
                    TGEV        PSGLVEPCIV RVSYGNNVLN GLWLGDEVIC PRHVIASDTT RV-INYENEM SSVRLHNFSV
                    OV43        PTSKVEPCVV SVTYGNMTLN GLWLDDKVYC PRHVICSASD MTNPDYTNLL CRVTSSDFTV
                    BoCoV       PTSKVEPCIV SVTYGNMTLN GLWLDDKVYC PRHVICSASD MTNPDYTNLL CRVTSSDFTV
                    MHV         PTSKVEPCVV SVTYGNMTLN GLWLDDKVYC PRHVICSSAD MTDPDYSNLL CRVISSDFCV
                    AIBV        PSSAVEKCIV SVSYRGNNLN GLWLGDTIYC PRHVLGKFSG DQ---WNDVL NLANNHEFEV
                    SARS CoV    PSGKVEGCMV QVTCGTTTLN GLWLDDTVYC PRHVICTAED MLNPNYEDLL IRKSNHSFLV

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3485       3495       3505       3515       3525       3535
                    EMCR        SHNG-VFLGV VGVTMHGSVL RIKVSQSNVH TPKHVFKTLK PGASFNILAC YEGIASGVFG
                    229E        ISGT-AFLGV VGATMHGVTL KIKVSQTNMH TPRHSFRTLK SGEGFNILAC YDGCAQGVFG
                    PEDV        SSGN-VFLGV VSATMRGALL QIKVNQNNVH TPKYTYRTVR PGESFNILAC YDGAAAGVYG
                    TGEV        SKNN-VFLGV VSARYKGVNL VLKVNQVNPN TPEHKFKSIK AGESFNILAC YEGCPGSVYG
                    OV43        LFDR-LSLTV MSYQMRGCML VLTVTLQNSR TPKYTFGVVK PGETFTVLAA YNGKPQGAFH
                    BoCoV       LFDR-LSLTV MSYQMQGCML VLTVTLQNSR TPKYTFGVVK PGETFTVLAA YNGKPQGAFH
                    MHV         MSGR-MSLTV MSYQMQGSLL VLTVTLQNPN TPKYSFGVVK PGETFTVLAA YNGKSQGAFH
                    AIBV        TTQHGVTLNV VSRRLKGAVL ILQTAVANAE TPKYKFIKAN CGDSFTIACA YGGTVVGLYP
                    SARS CoV    QAGN-VQLRV IGHSMQNCLL RLKVDTSNPK TPKYKFVRIQ PGQTFSVLAC YNGSPSGVYQ

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3545       3555       3565       3575       3585       3595
                    EMCR        VNLRTNFTIK GSFINGACGS PGYNVRNDGT VEFCYLHQIE LGSGAHVGSD FTGSVYGNFD
                    229E        VNMRTNWTIR GSFINGACGS PGYNLKN-GE VEFVYMHQIE LGSGSHVGSS FDGVMYGGFE
                    PEDV        VNMRSNYTIR GSFINGACGS PGYNINN-GT VEFCYLHQLE LGSGCHVGSD LDGVMYGGYE
                    TGEV        VNMRSQGTIK GSFIAGTCGS VGYVLEN-GI LYFVYMHHLE LGNGSHVGSN FEGEMYGGYE
                    OV43        VTMRSSYTIK GSFLCGSCGS VGYVIMG-DC VKFVYMHQLE LSTGCHTGTD FNGDFYGPYK
                    BoCoV       VTMRSSYTIK GSFLCGSCGS VGYVIMG-DC VKFVYMHQLE LSTGCHTGTD FNGDFYGPYK
                    MHV         VTMRSSYTIK GSFLCGSCGS VGYVLTG-DS VRFVYMHQLE LSTGCHTGTD FSGNFYGPYR
                    AIBV        VTMRSNGTIR ASFLAGACGS VGFNIEK-GV VNFFYMHHLE LPNALHTGTD LMGEFYGGYV
                    SARS CoV    CAMRPNHTIK GSFLNGSCGS VGFNIDY-DC VSFCYMHHME LPTGVHAGTD LEGKFYGPFV

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3605       3615       3625       3635       3645       3655
                    EMCR        DQPSLQVESA NLMLSDNVVA FLYAALLNGC R------WWL RSTRVNVDGF NEWAMANGYT
                    229E        DQPNLQVESA NQMLTVNVVA FLYAAILNGC T------WWL KGEKLFVEHY NEWAQANGFT
                    PEDV        DQPTLQVEGA SSLFTENVLA FLYAALINGS T------WWL SSSRIAVDRF NEWAVHNGMT
                    TGEV        DQPSMQLEGT NVMSSDNVVA FLYAALINGE R------WFV TNTSMSLESY NTWAKTNSFT
                    OV43        DAQVVQLLIQ DYIQSVNFVA WLYAAILNNC N------WFV QSDKCSVEDF NVWALSNGFS
                    BoCoV       DAQVVQLPVQ DYIQSVNFVA WLYAAILNNC N------WFV QSDKCSVEDF NVWALSNGFS
                    MHV         DAQVVQLPVQ DYTQTVNVVA WLYAAILNRC N------WFV QSDSCSLEEF NVWAMTNGFS
                    AIBV        DEEVAQRVPP DNLVTNNIVA WLYAAIISVK ESSFSLPKWL ESTTVSVDDY NKWAGDNGFT
                    SARS CoV    DRQTAQAAGT DTTITLNVLA WLYAAVINGD R------WFL NRFTTTLNDF NLVAMKYNYE

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3665       3675       3685       3695       3705       3715
                    EMCR        IVSSVEC--Y SILAAKTGVS VEQLLASIQH LHE-GFGGKN ILGYSSLCDE FTLAEVVKQM
                    229E        AMNGEDA--F SILAAKTGVC VERLLHAIQV LNN-GFGGKQ ILGYSSLNDE FSINEVVKQM
                    PEDV        TVGNTDC--F SILAAKTGVD VQRLLASIQS LHK-NFGGKQ ILGHTSLTDE FTTGEVVRQM
                    TGEV        ELSSTDA--F SMLAAKTGQS VEKLLDSIVR LNK-GFGGRT ILSYGSLCDE FTPTEVIRQM
                    OV43        QVKSDLV--I DALASMTGVS LETLLAAIKR LKN-GFQGRQ IMGSCSFEDE LTPSDVYQQL
                    BoCoV       QVKSDLV--I DALASMTGVS LETLLAAIKR LKN-GFQGRQ IMGSCSFEDE LTPSDVYQQL
                    MHV         SIKADLV--L DALASMTGVT VEQILAAIKR LYS-GFQGKQ ILGSCVLEDE LTPSDVYQQL
                    AIBV        PFSTSTA--I TKLSAITGVD VCKLLRTIMV KNS-QWGGDP ILGQYNFEDE LTPESVFNQI
                    SARS CoV    PLTQDHVDIL GPLSAQTGIA VLDMCAALKE LLQNGMNGRT ILGSTILEDE FTPFDVVRQC

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    3725       3735       3745       3755       3765       3775
```

```
EMCR        YGVNLQS--- GKVIFGLKTM FLFSVFFTMF WAELFIYTNT IWINPVILTP IFCLLLFLSL
229E        FGVNLQS--- GKTTSMFKSI SLFAGFFVMF WAELFVYTTT IWVNPGFLTP FMILLVALSL
PEDV        YGVNLQG--- GYVSRACRNV LLVGSFLTFF WSELVSYTKF FWVNPGYVTP MFACLSLLSS
TGEV        YGVNLQA--- GKVKSFFYPI MTAMTILFAF WLEFFMYTPF TWINPTFVSI VLAVTTLIST
OV43        AGIKLQSKRT RLFKGTVCWI MASTFLFSCI ITAFVKWTMF MYVTTNMFS- ITFCALCVIS
BoCoV       AGIKLQSKRT RLVKGIVCWI MASTFLFSCI ITAFVKWTMF MYVTTNMLS- ITFCALCVIS
MHV         AGVKLQSKRT RVVKGTCCWI LASTLLFCSI ISAFVKWTMF MYVTTHMLG- VTLCALCFVS
AIBV        GGVRLQS--- -SFVRKATSW FWSRCVLACF LFVLCAIVLF TAVPLKFYVY AAVILLMAVL
SARS CoV    SGVTFQGKFK KIVKGTHHWM LLTFLTSLLI LVQSTQWSLF FFVYENAFLP FTLGIMAIAA

                    3785      3795      3805      3815      3825      3835
EMCR        VLTMFLKHKF LFLQVFLLPT VIATALYN-- --CVLDYYIV KFLADHFN-Y NVSVLQMDVQ
229E        CLTFVVKHKV LFLQVFLLPS IIVAAIQN-- --CAWDYHVT KVLAEKFD-Y NVSVMQMDIQ
PEDV        LLMFTLKHKT LFFQVFLIPA LIVTSCIN-- --LAFDVEVY NYLAEHFD-Y HVSLMGFNAQ
TGEV        VFVSGIKHKM LFFMSFVLPS VILVTAHN-- --LFWDFSYY ESLQSIVENT NTMFLPVDMQ
OV43        LAMLLVKHKH LYLTMYITPV LFTLLYNNY- -LVVYKHTFR GYVYAWLSYY VPSVEYTYTD
BoCoV       LAMLLVKHKH LYLTMYIIPV LFTLLYNNY- -LVVYKQTFR GYVYAWLSYY VPSVEYTYTD
MHV         FAMLLVKHKH LYLTMFIMPV LCTLFYTNY- -LVVYKQSFR GLAYAWLSHF VPAVDYTYMD
AIBV        FISFTVKHVM AYMDTFLLPT LITVIIGVCA EVPFIYNTLI SQVVIFLSQW YDPVVFDTMV
SARS CoV    CAMLLVKHKH AFLCLFLLPS LATVAYFN-- -MVYMPASWV MRIMTWLELA DTSLSGYRLK

                    3845      3855      3865      3875      3885      3895
EMCR        GLVNVLVCLF VVFLH---TW RFSKERFTHW FTYVCSLIAV AYTYFYSGD- ----------F
229E        GFVNIFICLF VALLH---TW RFAKERCTHW CTYLFSLIAV LYTALYSYD- ----------Y
PEDV        GLVNIFVCFV VTILHGTYTW RFFN-TPASS VTYVVALLTA AYNYFYASD- ----------I
TGEV        GVMLTVFCFI VFVTYSVRFF TCKQSWFSLA VTTILVIFNM VKIFGTSDEP WTENQIAFCF
OV43        EVIYGMLLLV GMVFVTLRSI NHDLFSFIMF VGRLISVFSL WYKGSNLEEE ----------I
BoCoV       EVIYGMLLLI GMVFVTLRSI NHDLFSFIMF VGRVISVVSL WYMGSNLEEE ----------I
MHV         EVLYGVVLLV AMVFVTMRSI NHDVFSVMFL VGRLVSLVSM WYFGANLEEE ----------V
AIBV        PWMFLPLVLY TAFKCVQGCY MNSFNTSLLM LYQFVKLGFV IYTSSNTLTA YTEGNWELFF
SARS CoV    DCVMYASALV LLILMTARTV YDDAARRVWT LMNVITLVYK VYYGNALDQA ----------I

                    3905      3915      3925      3935      3945      3955
EMCR        LSLLVMFLCA ISSDWYIGAI VFRLSRLIIF FSPE--SVFS VFGDVKLTLV VYLICGYLVC
229E        VSLLVMLLCA ISNEWYIGAI IFRICRFGVA FLPV--EYVS YFDGVKTVLL FYMLLGFVSC
PEDV        LSCAMTLFAS VTGNWFVGAV CYKVAVYMAL RFP---TFVA IFGDIKSVMF CYLVLGYFTC
TGEV        VNMLTMIVSL TTKDWMVVIA SYRIAYYIVV CVMP-SAFVS DFGFMKCISI VYMACGYLFC
OV43        LLMLASLFGT YTWT---TVL SMAVAKVIAK WVAVNVLYFT DIPQIKIVLL CYLFIGYIIS
BoCoV       LLMLASLFGT YTWT---TAL SMAAAKVIAK WVAVNVLYFT DIPQIKIVLV CYLFIGYIIS
MHV         LLFLTSLFGT YTWT---TML SLATAKVIAK WLAVNVLYFT DVPQVKLVLL SYLCIGYVCC
AIBV        ELVHTTVLAN VSSNSLIGLF VFKCAKWMLY YCN-----AT YLNNYVLMAV MVNCIGWLCT
SARS CoV    SMWALVISVT SNYSGVVTTI MFLARAIVFV CVEYYPLLFI TGNTLQCIML VYCFLGYCCC

                    3965      3975      3985      3995      4005      4015
EMCR        TYWGILYWFN RFFKCTMGVY DFKVSAAEFK YMVANGLHAP YGPFDALWLS FKLLGIGGDR
229E        MYYGLLYWIN RFCKCTLGVY DFCVSPAEFK YMVANGLNAP NGPFDALFLS FKLMGIGGPR
PEDV        CFYGILYWFN RFFKVSVGVY DYTVSAAEFK YMVANGLRAP TGTLDSLLLS AKLIGIGGER
TGEV        CYYGILYWVN RFTCMTCGVY QFTVSAAELK YMTANNLSAP KNAYDAMILS AKLIGVGGKR
OV43        CYWGLFSLMN SLFRMPLGVY NYKISVQELR YMNANGLRPP KNSFEALMLN FKLLGIGGVP
BoCoV       CYWGLFSLMN SLFRMPLGVY NYKISVQELR YMNANGLRPP KNSFEALMLN FKLLGIGGVP
MHV         CYWGVLSLLN SIFRMPLGVY NYKISVQELR YMNANGLRPP RNSFEALVLN FKLLGIGGVP
AIBV        CYFGLYWWVN KVFGLTLGKY NFKVSVDQYR YMCLHKINPP KTVWEVFSTN ILIQGIGGDR
SARS CoV    CYFGLFCLLN RYFRLTLGVY DYLVSTQEFR YMNSQGLLPP KSSIDAFKLN IKLLGIGGKP

                    4025      4035      4045      4055      4065      4075
EMCR        CIKISTVQSK LTDLKCTNVV LLGCLSSMNI AANSSEWAYC VDLHNKINLC DDPEKAQGML
229E        TIKVSTVQSK LTDLKCTNVV LMGILSNMNI ASNSKEWAYC VEMHNKINLC DDPETAQELL
PEDV        NIKISSVQSK LTDIKCSNVV LLGCLSSMNV SANSTEWAYC VDLHNKINLC NDPEKAQEML
TGEV        NIKISTVQSK LTEMKCTNVV LLGLLSKMHV ESNSKEWNYC VGLHNEINLC DDPEIVLEKL
OV43        IIEVSQFQSK LTDVKCANVV LLNCLQHLHV ASNSKLWHYC STLHNEILAT SDLSVAFEKL
BoCoV       IIEVSQFQSK LTDVKCANGG LLNCLQHLHV ASNSKLWQYC STLHNEILAT SDLGVAFEKL
MHV         VIEVSQIQSR LTDVKCVNVV LLNCLQHLHI ASSSKLWQYC STLHNEILAT SDLSVAFDKL
AIBV        VLPIATVQAK LSDVKCTTVV LMQLLTKLNV EANSKMHVYL VELHNKILAS DDVGECMDNL
SARS CoV    CIKVATVQSK MSDVKCTSVV LLSVLQQLRV ESSSKLWAQC VQLHNDILLA KDTTEAFEKM

                    4085      4095      4105      4115      4125      4135
EMCR        LALLAFFLSK HSDFG----- -LDGLIDSYF DNSSTLQSVA SSFVSMPSYI AYENARQAYE
229E        LALLAFFLSK HSDFG----- -LGDLVDSYF ENDSILQSVA SSFVGMPSFV AYETARQEYE
PEDV        LALLAFFLSK NSAFG----- -LDDLLESYF NDNSMLQSVA STYVGLPSYV IYENARQQYE
TGEV        LALIAFFLSK HNTCD----- -LSELIESYF ENTTILQSVA SAYAALPSWI ALEKARADLE
OV43        AQLLIVLFAN PAAVDSKCLT SIEEVCDDYA KDNTVLQALQ SEFVNMASFV EYEVAKKNLD
BoCoV       AQLLIVLFAN PAAVDSKCLT SIEEVCDDYA KDNTVLQALQ SEFVNMASFV EYEVAKKNLD
MHV         AQLLVVLFAN PAAVDSKCLA SIEEVSDDYV RDSTVLQALQ SEFVNMASFV EYELAKKNLD
AIBV        LGMLITLFCI DSTID----- -LSEYCDDIL KRSTVLQSVT QEFSHIPSYA EYERAKNLYE
```

```
                SARS CoV    VSLLSVLLSM QGAVD----- -INRLCEEML DNRATLQAIA SEFSSLPSYA AYATAQEAYE

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4145       4155       4165       4175       4185       4195
                EMCR        DAIANGSS-- --SQLIKQLK RAMNIAKSEF DHEISVQKKI NRMAEQAATQ MYKEARSVNR
                229E        NAVANGSS-- --PQIIKQLK KAMNVAKAEF DRESSVQKKI NRMAEQAAAA MYKEARAVNR
                PEDV        DAVNNGSP-- --PQLVKQLR HAMNVAKSEF DREASTQRKL DRMAEQAAAQ MYKEARAVNR
                TGEV        EAKKNDVS-- --PQILKQLT KAFNIAKSDF EREASVQKKL DKMAEQAAAS MYKEARAVDR
                OV43        EARFSGSAN- --QQQLKQLE KACNIAKSAY ERDRAVAKKL ERMADLALTN MYKEARINDK
                BoCoV       EACSSGSAN- --QQQLKQLE KACNIAKSAY ERDRAVARKL ERMADLALTN MYKEARINDK
                MHV         EAKASGSAN- --QQQIKQLE KACNIAKSAY ERDRAVARKL ERMADLALTN MYKEARINDK
                AIBV        KVLVDSKNGG VTQQELAAYR KAANIAKSVF DRDLAVQKKL DSMAERAMTT MYKEARVTDR
                SARS CoV    QAVANGDS-- --EVVLKKLK KSLNVAKSEF DRDAAMQRKL EKMADQAMTQ MYKQARSEDK

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4205       4215       4225       4235       4245       4255
                EMCR        KSKVISAMHS LLFGMLRRLD MSSVETVLNL ARDGVVPLSV IPATSASKLT IVSPDLESYS
                229E        KSKVVSAMHS LLFGMLRRLD MSSVDTILNM ARNGVVPLSV IPATSAARLV VVVPDHDSFV
                PEDV        KSKVVSAMHS LLFGMLRRLD MSSVDTILNL AKDGVVPLSV IPAVSATKLN IVTSDIDSYN
                TGEV        KSKIVSAMHS LLFGMLKKLD MSSVNTIIDQ ARNGVLPLSI IPAASATRLV VITPSLEVFS
                OV43        KSKVVSALQT MLFSMVRKLD NQALNSILDN AVKGCVPLNA IPSLAANTLN IIVPDKSVYD
                BoCoV       KSKVVSALQT MLFSMVRKLD NQALNSILDN AVKGCVPLNA IPSLAANTLN IIVPDKSVYD
                MHV         KSKVVSALQT MLFSMIRKLD NQALNSILDN AVKGCVPLNA IPSLAANTLT IIVPDKSVYD
                AIBV        RAKLVSSLHA LLFSMLKKID SEKLNVLFDQ ASSGVVPLAT VPIVCSNKLT LVIPDPETWV
                SARS CoV    RAKVTSAMQT MLFTMLRKLD NDALNNIINN ARDGCVPLNI IPLTTAAKLM VVVPDYGTYK

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4265       4275       4285       4295       4305       4315
                EMCR        KIVCDGSVHY AGVVWTLNDV KDNDGRPVHV KEITRENVET LT-----WPL ILNCERVVK-
                229E        KMMVDGFVHY AGVVWTLQEV KDNDGKNVHL KDVTKENQEI LV-----WPL ILTCERVVK-
                PEDV        RIQREGCVHY AGTIWNIIDI KDNDGKVVHV KEVTAQNAES LS-----WPL VLGCERIVK-
                TGEV        KIRQENNVHY AGAIWTIVEV KDANGSHVHL KEVTAANELN LT-----WPL SITCERTTK-
                OV43        QVVDNVYVTY AGNVWQIQTI QDSDGTNKQL NEISDDCN-- -------WPL VIIANRYNE-
                BoCoV       QVVDNVYVTY AGNVWQIQTI QDSDGTNKQL HEISDDCN-- -------WPL VIIANRYNE-
                MHV         QVVDNVYVTY AGNVWHIQSI QDADGAVKQL NEIDVNIT-- -------WPL VIAANRHNE-
                AIBV        KCVEGVHVTY STVVWNIDTV IDADGTELHP TSTGSGLTYC ISGANIAWPL KVNLTRNGHN
                SARS CoV    NTCDGNTFTY ASALWEIQQV VDADSKIVQL SEINMDNSPN LA-----WPL IVTALRAN--

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4325       4335       4345       4355       4365       4375
                EMCR        ----LQ-NNE IMPGKLKQKP MKAEG--DGG VLGDGNALYN TEGGKTFMYA YISNKADLKF
                229E        ----LQ-NNE IMPGKMKVKA TKGEG--DGG ITSEGNALYN NEGGRAFMYA YVTTKPGMKY
                PEDV        ----LQ-NNE IIPGKLKQRS IKAEG--DG- IVGEGKALYN NEGGRTFMYA FISDKPDLRV
                TGEV        ----LQ-NNE IMPGKLKERA VRASATLDGE AFGSGKALMA SESGKSFMYA FIASDNNLKY
                OV43        VSATVLQNNE LMPAKLKIQV VNSGPDQTCN TPT--QCYYN NSNNGKIVYA ILSDVDGLKY
                BoCoV       VSATVLQNNE LMPAKLKTQV VNSGPDQTCN TPT--QCYYN NSYNGKIVYA ILSDVDGLKY
                MHV         VSSVVLQNNE LMPQKLRTQV VNSGSDMNCN TPT--QCYYN TTGMGKIVYA ILSDCDGLKY
                AIBV        KVDVVLQNNE LMPHGVKTKA CVAGVDQAHC SVES-KCYYT NISGNSVVAA ITSSNPNLKV
                SARS CoV    -SAVKLQNNE LSPVALRQMS CAAGTTQTAC TDDNALAYYN NSKGGRFVLA LLSDHQDLKW

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4385       4395       4405       4415       4425       4435
                EMCR        VKWEYEGG-- CNTIELDSPC RFMVETPNGP QVKYLYFVKN LNTLRRGAVL GFIGATIRLQ
                229E        VKWEHDSG-- VVTVELEPPC RFVIDTPTGP QIKYLYFVKN LNNLRRGAVL GYIGATVRLQ
                PEDV        VKWEFDGG-- CNTIELEPPR KFLVDSPNGA QIKYLYFVRN LNTLRRGAVL GYIGATVRLQ
                TGEV        VKWESNND-- IIPIELEAPL RFYVDGANGP EVKYLYFVKN LNTLRRGAVL GYIGATVRLQ
                OV43        TKILKDDGN- FVVLELDPPC KFTVQDAKGL KIKYLYFVKG CNTLARGWVV GTISSTVRLQ
                BoCoV       TKILKDDGN- FVVLELDPPC KFTVQDVKGL KIKYLYFVKG CNTLARGWVV GTISSTVRLQ
                MHV         TKIVKEDGN- CVVLELDPPC KFSVQDVKGL KIKYLYFVKG CNTLARGWVV GTISSTVRLQ
                AIBV        ASFLNEAGN- QIYVDLDPPC KFGMKVGVKV EVVYLYFIKN TRSIVRGMVL GAISNVVVLQ
                SARS CoV    ARFPKSDGTG TIYTELEPPC RFVTDTPKGP KVKYLYFIKG LNNLNRGMVL GSLAATVRLQ

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4445       4455       4465       4475       4485       4495
                EMCR        AG-KQTELAV NSGLLTACAF SVDPATTYLE AVKHGAKPVS NCIKMLSNGA GNGQAITTSV
                229E        AG-KQTEFVS NSHLLTHCSF AVDPAAAYLD AVKQGAKPVG NCVKMLTNGS GSGQAITCTI
                PEDV        AG-KQTEQAI NSSLLTLCAF AVDPAKTYID AVKSGHKPVG NCVKMLANGS GNGQAVTNGV
                TGEV        AG-KPTEHPS NSSLLTLCAF SPDPAKAYVD AVKRGMQPVN NCVKMLSNGA GNGMAVTNGV
                OV43        AG-TATEYAS NSSILSLCAF SVDPKKTYLD FIQQGGTPIA NCVKMLCDHA GTGMAITVKP
                BoCoV       AG-TATEYAS NSSILSLCAF SVDPKKTYLD FIQQGGTPIA NCVKMLCDHA GTGMAITVKP
                MHV         AG-TATEYAS NSAIRSLCAF SVDPKKTYLD YIQQGGAPVT NCVKMLCDHA GTGMAITIKP
                AIBV        SKGHETEEVD AVGILSLCSF AVDPADTYCK YVAAGNQPLG NCVKMLTVHN GSGFAITSKP
                SARS CoV    AG-NATEVPA NSTVLSFCAF AVDPAKAYKD YLASGGQPIT NCVKMLCTHT GTGQAITVTP

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              4505       4515       4525       4535       4545       4555
                EMCR        DANTNQDSYG GASICLYCRA HVPHP----S MDGYCKFKGK CVQVP-IGCL DPIRFCLENN
                229E        DSNTTQDTYG GASVCIYCRA HVAHP----T MDGFCQYKGK WVQVP-IGTN DPIRFCLENT
                PEDV        EASTNQDSYG GASVCLYCRA HVEHP----S MDGFCRLKGK YVQVP-LGTV DPIRFVLEND
                TGEV        EANTQQDSYG GASVCIYCRC HVEHP----A IDGLCRYKGK FVQIP-TGTQ DPIRFCIENE
```

```
     OV43        DATTSQDSYG GASVCIYCRA RVEHP----D VDGLCKLRGK FVQVP-VGIK DPVSYVLTHD
     BoCoV       DATTSQDSYG GASVCIYCRA RVEHP----D VDGLCKLRGK FVQVP-VGIK DPVSYVLTHD
     MHV         EATTNQDSYG GASVCIYCRS RVEHP----D VDGLCKLRGK FVQVP-LGIK DPVSYVLTHD
     AIBV        SPTPDQDSYG GASVCLYCRA HIAHPGSVGN LDGRCQFKGS FVQIP-TTEK DPVGFCLRNK
5    SARS CoV    EANMDQESFG GASCCLYCRC HIDHP----N PKGFCDLKGK YVQIPTTCAN DPVGFTLRNT

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4565       4575       4585       4595       4605       4615
     EMCR        VCNVCGCWLG HGCACDRTTI QSVDIS---- YLNRARGSSA -ARLEPCN-G TDIDKCVRAF
10   229E        VCKVCGCWLN HGCTCDRTAI QSFDNS---- YLNRVRGSSA -ARLEPCN-G TDIDYCVRAF
     PEDV        VCKVCGCWLS NGCTCDRSIM QSTDYG---- LFKRVRGSSA -ARLEPCN-G TDTQHVYRAF
     TGEV        VCVVCGCWLN NGCMCDRTSM QSFTVDQSY- LFKRVRGSSA -ARLEPCN-G TDPDHVSRAF
     OV43        VCRVCGFWRD GSCSCVSTDT TVQSKDTN-- FFKRVRGTSV DARLVPCASG LSTDVQLRAF
     BoCoV       VCQVCGFWRD GSCSCVSTDT TVQSKDT--- FFKRVRGTSV DARLVPCASG LSTDVQLRAF
15   MHV         VCQVCGFWRD GMFLCR-HRL PVSVKRHE-- LFKRVRGTSV NARLVPCASG LDTDVQLRAF
     AIBV        VCTVCQCWIG YGCQCDSLRQ PKSSVQS--- ----VAGASD FDKNYLNG-- YGVAVRLGMF
     SARS CoV    VCTVCGMWKG YGCSCDQLRE PLMQSADAST FLNRVCGVSA -ARLTPCGTG TSTDVVYRAF

20               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4625       4635       4645       4655       4665       4675
     EMCR        DIYNKNVSFL GKCLKMNCVR FKNADLK--- -DGYFVIKRC TKSVMEHEQS MYNLLNFSGA
     229E        DVYNKDASFI GKNLKSNCVR FKNVDKD--- -DAFYIVKRC IKSVMDHEQS MYNLLKGCNA
     PEDV        DIYNKDVACL GKFLKVNCVR LKNLDKH--- -DAFYVVKRC TKSAMEHEQS IYSRLEKCGA
     TGEV        DIYNKDVACI GKFLKTNCSR FRNLDKH--- -DAYYIVKRC TKTVMDHEQV CYNDLKDSGA
25   OV43        DIYNASVAGI GLHLKVNCCR FQRVDENGDK LDQFFVVKRT DLTIYNREMK CYERVKDCKF
     BoCoV       DICNASVAGI GLHLKVNCCR FQRVDENGDK LDQFFVVKRT DLTIYNREME CYERVKDCKF
     MHV         DICNANRAGI GLYYKVNCCR FQRADEDGNT LDKFFVIKRT NLEVYNKEKE CYELTKECGV
     AIBV        QNLKRNCARF QELRDTEDGN LEYLDS---- ---YFVVKQT TPSNYEHEKS CYEDLKS-EV
     SARS CoV    DIYNEKVAGF AKFLKTNCCR FQEKDEEGNL LDSYFVVKRH TMSNYQHEET IYNLVKDCPA
30
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4685       4695       4705       4715       4725       4735
     EMCR        LAEHDFFTWK DGRVIYGNVS RHNLTKYTMM DLVYAMRNFD EQNCDVLKEV LVLTGCCDNS
     229E        VAKHDFFTWH EGRTIYGNVS RQDLTKYTMM DLCFALRNFD EKDCEVFKEI LVLTGCCSTD
35   PEDV        IAEHDFFTWK DGRAIYGNVC RKDLTEYTMM DLCYALRNFD ENNCDVLKSI LIKVGACEES
     TGEV        VAEHDFFTYK EGRCEFGNVA RRNLTKYTMM DLCYAIRNFD EKNCEVLKEI LVTVGACTEE
     OV43        VAEHDFFTFD VEGSRVPHIV RKDLTKYTML DLCYALRHFD RNDCMLLCDI LSIYAGCEQS
     BoCoV       VAEHDFFTFD VEGSRVPHIV RKDLTKYTML DLCYALRHFD RNDCMLLCDI LSIYAGCEQS
     MHV         VAEHEFFTFD VEGSRVPHIV RKDLSKYTML DLCYALRHFD RNDCSTLKEI LLTYAECDES
40   AIBV        TADHDFFVFN KN---IYNIS RQRLTKYTMM DFCYALRHFD PKDCEVLKEI LVTYGCIEDY
     SARS CoV    VAVHDFFKFR VDGDMVPHIS RQRLTKYTMA DLVYALRHFD EGNCDTLKEI LVTYNCCDDD

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4745       4755       4765       4775       4785       4795
45   EMCR        ----YFDSKG WYDPVENEDI HRVYASLGKI VARAMLKCVA LCDAMVAKGV VGVLTLDNQD
     229E        ----YFEMKN WFDPIENEDI HRVYAALGKV VANAMLKCVA FCDEMVLKGV VGVLTLDNQD
     PEDV        ----YFNNKV WFDPVENEDI HRVYALLGTI VARAMLKCVK FCDAMVEQGI VGVVTLDNQD
     TGEV        ----FFENKD WFDPVENEAI HEVYAKLGPI VANAMLKCVA FCDAIVEKGY IGVITLDNQD
     OV43        ----YFTKKD WYDFVENPDI INVYKKLGPI FNRALVSATE FADKLVEVGL VGVLTLDNQD
50   BoCoV       ----YFTKKD WYDFVENPDI INVYKKLGPI FNRALVSATE FADKLVEVGL VGILTLDNQD
     MHV         ----YFQKKD WYDFVENSDI INVYKKLGPI FNRALLNTAK FADTLVEAGL VGVLTLDNQD
     AIBV        HPKWFEENKD WYDPIENSKY YVMLAKMGPI VRRALLNAIE FGNLMVEKGY VGVITLDNQD
     SARS CoV    ----YFNKKD WYDFVENPDI LRVYANLGER VRQSLLKTVQ FCDAMRDAGI VGVLTLDNQD

55               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4805       4815       4825       4835       4845       4855
     EMCR        LNGNFYDFGD FVVSLPNMGV PCCTSYYSYM MPIMGLTNCL ASECFVKSDI FGSDFKTFDL
     229E        LNGNFYDFGD FVLCPPGMGI PYCTSYYSYM MPVMGMTNCL ASECFMKSDI FGQDFKTFDL
     PEDV        LNGDFYDFGD FTCSIKGMGV PICTSYYSYM MPVMGMTNCL ASECFVKSDI FGEDFKSYDL
60   TGEV        LNGNFYDFGD FVKTAPGFGC ACVTSYYSYM MPLMGMTSCL ESENFVKSDI YGSDYKQYDL
     OV43        LNGKWYDFGD YVIAAPGCGV AIADSYYSYI MPMLTMCHAL DCELYVNN-- ---AYRLFDL
     BoCoV       LNGKWYDFGD YVIAAPGCGV AIADSYYSYM MPMLTMCHAL DCELYVNN-- ---AYRLFDL
     MHV         LYGQWYDFGD FVKTVPGCGV AVADSYYSYM MPMLTMCHAL DSELFING-- ---TYREFDL
     AIBV        LNGKFYDFGD FQKTAPGAGV PVFDTYYSYM MPIIAMTDAL APERYFEYDV H-KGYKSYDL
65   SARS CoV    LNGNWYDFGD FVQVAPGCGV PIVDSYYSLL MPILTLTRAL AAESHMDADL A-KPLIKWDL

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4865       4875       4885       4895       4905       4915
     EMCR        LKYDFTEHKE NLFNKYFKHW SFDYHPNCSD CYDDMCVIHC ANFNTLFATT IPGTAFGPLC
70   229E        LKYDFTEHKE VLFNKYFKYW GQDYHPDCVD CHDEMCILHC SNFNTLFATT IPNTAFGPLC
     PEDV        LEYDFTEHKT ALFNKYFKYW GLQYHPNCVD CSDEQCIVHC ANFNTLFSTT IPITAFGPLC
     TGEV        LAYDFTEHKE YLFQKYFKYW DRTYHPNCSD CTSDECIIHC ANFNTLFSMT IPMTAFGPLV
     OV43        VQYDFTDYKL ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV
     BoCoV       VQYDFTDYKL ELFNKYFKHW SMPYHPNTVD CQDDRCIIHC ANFNILFSMV LPNTCFGPLV
75   MHV         VQYDFTDFKL ELFNKYFKYW SMTYHPNTCE CEDDRCIIHC ANFNILFSMV LPKTCFGPLV
     AIBV        LKYDYTEEKQ ELFQKYFKYW DQEYHPNCRD CSDDRCLIHC ANFNILFSTL IPQTSFGNLC
     SARS CoV    LKYDFTEERL CLFDRYFKYW DQTYHPNCIN CLDDRCILHC ANFNVLFSTV FPPTSFGPLV

80               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   4925       4935       4945       4955       4965       4975
```

```
          EMCR       RKVFIDGVPL VTTAGYHFKQ LGLVWNKDVN THSVRLTITE LLQFVTDPSL IIASSPALVD
          229E       RKVFIDGVPV VATAGYHFKQ LGLVWNKDVN THSTRLTITE LLQFVTDPTL IVASSPALVD
          PEDV       RKCWIDGVPL VTTAGYHFKQ LGIVWNNDLN LHSSRLSINE LLQFCSDPAL LIASSPALVD
          TGEV       RKVHIDGVPV VVTAGYHFKQ LGIVWNLDVK LDTMKLSMTD LLRFVTDPTL LVASSPALLD
          OV43       RQIFVDGVPF VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD
          BoCoV      RQIFVDGVPF VVSIGYHYKE LGIVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD
          MHV        RQIFVDGVPF VVSIGYHYKE LGVVMNMDVD THRYRLSLKD LLLYAADPAL HVASASALYD
          AIBV       RKVFVDGVPF IATCGYHSKE LGVIMNQDNT MSFSKMGLSQ LMQFVGDPAL LVGTSNNLVD
          SARS CoV   RKIFVDGVPF VVSTGYHFRE LGVVHNQDVN LHSSRLSFKE LLVYAADPAM HAASGNLLLD

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          4985       4995       5005       5015       5025       5035
          EMCR       QRTICFSVAA LSTGLTNQVV KPGHFNEEFY NFLRLRGFFD EGSELTLKHF FFAQNGDAAV
          229E       KRTVCFSVAA LSTGLTSQTV KPGHFNKEFY DFLRSQGFFD EGSELTLKHF FFTQKGDAAI
          PEDV       QRTVCFSVAA LGTGMTNQTV KPGHFNKEFY DFLLEQGFFS EGSELTLKHF FFAQKVDAAV
          TGEV       QRTVCFSIAA LSTGITYQTV KPGHFNKDFY DFITERGFFE EGSELTLKHF FFAQGGEAAM
          OV43       LRTCCFSVAA ITSGVKFQTV KPGNFNQDFY DFVLSKGLLK EGSSVDLKHF FFTQDGNAAI
          BoCoV      LRTCCFSVAA ITSGVKFQTV KPGNFNQDFY DFILSKGLLK EGSSVDLKHF FFTQDGNAAI
          MHV        LRTCCFSVAA ITSGVKFQTV KPGNFNQDFY EFILSKGLLK EGSSVDLKHF FFTQDGNAAI
          AIBV       LRTSCFSVCA LTSGITHQTV KPGHFNKDFY DFAEKAGMFK EGSSIPLKHF FYPQTGNAAI
          SARS CoV   KRTTCFSVAA LTNNVAFQTV KPGNFNKDFY DFAVSKGFFK EGSSVELKHF FFAQDGNAAI

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          5045       5055       5065       5075       5085       5095
          EMCR       KDFDFYRYNK PTILDICQAR VTYKIVSRYF DIYEGGCIKA CEVVVTNLNK SAGWPLNKFG
          229E       KDFDYYRYNR PTMLDIGQAR VAYQVAARYF DCYEGGCITS REVVVTNLNK SAGWPLNKFG
          PEDV       KDFDYYRYNR PTVLDICQAR VVYQIVQRYF DIYEGGCITA KEVVVTNLNK SAGYPLNKFG
          TGEV       TDFNYYRYNR VTVLDICQAQ FVYKIVGKYF ECYDGGCINA REVVVTNYDK SAGYPLNKFG
          OV43       TDYNYYKYNL PTMVDIKQLL FVLEVVYKYF EIYDGGCIPA SQVIVNNYDK SAGYPLNKFG
          BoCoV      TDYNYYKYNL PTMVDIKQLL FVLEVVYKYF EIYDGGCIPA SQVIVNNYDK SAGYPFNKFG
          MHV        TDYNYYKYNL PTMVDIKQLL FVLEVVNKYF EIYDGGCIPA TQVIVNNYDK SAGYPFNKFG
          AIBV       NDYDYYRYNR PTMFDICQLL FCLEVTSKYF ECYEGGCIPA SQVVVNNLDK SAGYPFNKFG
          SARS CoV   SDYDYYRYNL PTMCDIRQLL FVVEVVDKYF DCYDGGCINA NQVIVNNLDK SAGFPFNKWG

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          5105       5115       5125       5135       5145       5155
          EMCR       KASLYYESIS YEEQDALFAL TKRNVLPTMT QLNLKYAISG KERARTVGGV SLLSTMTTRQ
          229E       KAGLYYESIS YEEQDAIFSL TKRNILPTMT QLNLKYAISG KERARTVGGV SLLATMTTRQ
          PEDV       KAGLYYESLS YEEQDELYAY TKRNILPTMT QLNLKYAISG KERARTVGGV SLLSTMTTRQ
          TGEV       KARLYYETLS YEEQDALFAL TKRNVLPTMT QMNLKYAISG KARARTVGGV SLLSTMTTRQ
          OV43       KARLYYEALS FEEQDEIYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM
          BoCoV      KARLYYEALS FEEQDEIYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM
          MHV        KARLYYEALS FEEQDEVYAY TKRNVLPTLT QMNLKYAISA KNRARTVAGV SILSTMTGRM
          AIBV       KARLYYEMS- LEEQDQLFEI TKKNVLPTIT QMNLKYAISA KNRARTVAGV SILSTMTNRQ
          SARS CoV   KARLYYDSMS YEDQDALFAY TKRNVIPTIT QMNLKYAISA KNRARTVAGV SICSTMTNRQ

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          5165       5175       5185       5195       5205       5215
          EMCR       YHQKHLKSIV NTRNATVVIG TTKFYGGWNN MLRTLIDGVE NPMLMGWDYP KCDRALPNMI
          229E       FHQKCLKSIV ATRNATVVIG TTKFYGGWDN MLKNLMADVD DPKLMGWDYP KCDRAMPSMI
          PEDV       YHQKHLKSIV NTRGASVVIG TTKFYGGWDN MLKNLIDGVE NPCLMGWDYP KCDRALPNMI
          TGEV       YHQKHLKSIA ATRNATVVIG STKFYGGWDN MLKNLMRDVD NGCLMGWDYP KCDRALPNMI
          OV43       FHQKCLKSIA ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD NPVLMGWDYP KCDRALPNMI
          BoCoV      FHQKCLKSIA ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD NPVLMGWDYP KCDRAMPNLL
          MHV        FHQKCLKSIA ATRGVPVVIG TTKFYGGWDD MLRRLIKDVD NPVLMGWDYP KCDRAMPNIL
          AIBV       FHQKILKSIV NTRNASVVIG TTKFYGGWDN MLRNLIQGVE DPILMGWDYP KCDRAMPNIL
          SARS CoV   FHQKLLKSIA ATRGATVVIG TSKFYGGWHN MLKTVYSDVE TPHLMGWDYP KCDRAMPNML

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          5225       5235       5245       5255       5265       5275
          EMCR       RMISAMVLGS KHVNCCTVTD RFYRLGNELA QVLTEVVYSN GGFYFKPGGT TSGDASTAYA
          229E       RMLSAMILGS KHVTCCTASD KFYRLSNELA QVLTEVVYSN GGFYFKPGGT TSGDATTAYA
          PEDV       RMISAMILGS KHTTCCSSTD RFFRLCNELA QVLTEVVYSN GGFYLKPGGT TSGDATTAYA
          TGEV       RMASAMILGS KHVGCCTHND RFYRLSNELA QVLTEVVHCT GGFYFKPGGT TSGDGTTAYA
          OV43       RIVSSLVLAR KHETCCSQSD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA
          BoCoV      RIVSSLVLAR KHEACCSQSD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA
          MHV        RIISSLVLAR KHDSCCSHTD RFYRLANECA QVLSEIVMCG GCYYVKPGGT SSGDATTAFA
          AIBV       RIAASLVLAR KHTNCCSWSE RIYRLYNECA QVLSETVLAT GGIYVKPGGT SSGDATTAYA
          SARS CoV   RIMASLVLAR KHNTCCNLSH RFYRLANECA QVLSEMVMCG GSLYVKPGGT SSGDATTAYA

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          5285       5295       5305       5315       5325       5335
          EMCR       NSIFNIFQAV SSNINRLLSV PSDSCNNVNV RDLQRRLYDN CYRLTSVEES FIDDYYGYLR
          229E       NSVFNIFQAV SSNINCVLSV NSSNCNNFNV KKLQRQLYDN CYRNSNVDES FVDDFYGYLQ
          PEDV       NSVFNIFQAV SANVNKLLSV DSNVCHNLEV KQLQRKLYEC CYRSTIVDDQ FVVEYYGYLR
          TGEV       NSAFNIFQAV SANVNKLLGV DSNACNNVTV KSIQRKIYDN CYRSSSIDEE FVVEYFSYLR
          OV43       NSVFNICQAV SANVCALMSC NGNKIEDLSI RALQKRLYSH VYRSDKVDST FVTEYYEFLN
          BoCoV      NSVFNICQAV SANVCALMSC NGNKIEDLSI RALQKRLYSH VYRSDMVDST FVTEYYEFLN
          MHV        NSVFNICQAV SANVCSLMAC NGHKIEDLSI RELQKRLYSN VYRADHVDPA FVNEYYEFLN
          AIBV       NSVFNIIQAT SANVARLLSV ITRDIVYDNI KSLQYELYQQ VYRRVNFDPA FVEKFYSYLC
```

```
SARS CoV    NSVFNICQAV TANVNALLST DGNKIADKYV RNLQHRLYEC LYRNRDVDHE FVDEFYAYLR

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5345       5355       5365       5375       5385       5395
EMCR        KHFSMMILSD DGVVCYNKDY AELGYIADIS AFKATLYYQN NVFMSTSKCW VEEDLTKGPH
229E        KHFSMMILSD DSVVCYNKTY AGLGYIADIS AFKATLYYQN GVFMSTAKCW TEEDLSIGPH
PEDV        KHFSMMILSD DGVVCYNNDY ASLGYVADLN AFKAVLYYQN NVFMSASKCW IEPDINKGPH
TGEV        KHFSMMILSD DGVVCYNKDY ADLGYVADIN AFKATLYYQN NVFMSTSKCW VEPDLSVGPH
OV43        KHFSMMILSD DGVVCYNSDY ASKGYIANIS AFQQVLYYQN NVFMSESKCW VEHDINNGPH
BoCoV       KHFSMMILSD DGVVCYNSDY ASKGYIANIS AFQQVLYYQN NVFMSESKCW VENDINNGPH
MHV         KHFSMMILSD DGVVCYNSEF ASKGYIANIS AFQQVLYYQN NVFMSEAKCW VETDIEKGPH
AIBV        KNFSLMILSD DGVVCYNNTL AKQGLVADIS GFREVLYYQN NVFMADSKCW VEPDLEKGPH
SARS CoV    KHFSMMILSD DAVVCYNSNY AAQGLVASIK NFKAVLYYQN NVFMSEAKCW TETDLTKGPH

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5405       5415       5425       5435       5445       5455
EMCR        EFCSQHTMQI VDKDGTYYLP YPDPSRILSA GVFVDDVVKT DAVVLLKRYV SLAIDAYPLS
229E        EFCSQHTMQI VDENGKYYLP YPDPSRIISA GVFVDDITKT DAVILLERYV SLAIDAYPLS
PEDV        EFCSQHTMQI VDKEGTYYLP YPDPSRILSA GVFVDDVVKT DAVVLLERYV SLAIDAYPLS
TGEV        EFCSQHTLQI VGPDGDYYLP YPDPSRILSA GVFVDDIVKT DNVIMLERYV SLAIDAYPLT
OV43        EFCSQHTMLV KMDGDDVYLP YPNPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV
BoCoV       EFCSQHTMLV KMDGDDVYLP YPVPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV
MHV         EFCSQHTMLV KMDGDEVYLP YPDPSRILGA GCFVDDLLKT DSVLLIERFV SLAIDAYPLV
AIBV        EFCSQHTMLV EVDGEPKYLP YPDPSRILGA CVFVDDVDKT EPVAVMERYI ALAIDAYPLV
SARS CoV    EFCSQHTMLV KQGDDYVYLP YPDPSRILGA GCFVDDIVKT DGTLMIERFV SLAIDAYPLT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5465       5475       5485       5495       5505       5515
EMCR        KHPNSEYRKV FYVLLDWVKH LNKNLNEGVL ESFSVTLLDN QEDKFWCEDF YASMYENSTI
229E        KHPKPEYRKV FYALLDWVKH LNKTLNEGVL ESFSVTLLDE HESKFWDESF YASMYEKSTV
PEDV        KHENPEYKKV FYVLLDWVKH LYKTLNAGVL ESFSVTLLED STAKFWDESF YANMYEKSAV
TGEV        KHPKPAYQKV FYTLLDWVKH LQKNLNAGVL DSFSVTMLEE GQDKFWSEEF YASLYEKSTV
OV43        YHENEEYQKV FRVYLAYIKK LYNDLGNQIL DSYSVILSTC DGQKFTDESF YKNMYLRSAV
BoCoV       YHENEEYQKV FRVYLEYIKK LYNELGNQIL DSYSVILSTC DGQKFTDESF YKNMYLRSAV
MHV         YHENPEYQNV FRVYLEYIKK LYNDLGNQIL DSYSVILSTC DGQKFTDETF YKNMYLRSAV
AIBV        HHENEEYKKV FFVLLAYIRK LYQELSQNML MDYSFVMDID KGSKFWEQEF YENMYRAPTT
SARS CoV    KHPNQEYADV FHLYLQYIRK LHDELTGHML DMYSVMLTND NTSRYWEPEF YEAMYTPHTV

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5525       5535       5545       5555       5565       5575
EMCR        LQAAGLCVVC GSQTVLRCGD CLRKPMLCTK CAYDHVFGTD HKFILAITPY VCNASGCGVS
229E        LQAAGLCVVC GSQTVLRCGD CLRRPMLCTK CAYDHVFGTD HKFILAITPY VCNTSGCNVN
PEDV        LQSAGLCVVC GSQTVLRCGD CLRRPMLCTK CAYDHVIGTT HKFILAITPY VCCASDCGVN
TGEV        LQAAGMCVVC GSQTVLRCGD CLRRPLLCTK CAYDHVMGTK HKFIMSITPY VCSFNGCNVN
OV43        MQSVGACVVC SSQTSLRCGS CIRKPLLCCK CCYDHVMATD HKYVLSVSPY VCNAPGCDVN
BoCoV       MQSVGACVVC SSQTSLRCGS CIRKPLLCCK CCYDHVMATD HKYVLSVSPY VCNAPGCDVN
MHV         MQSVGACVVC SSQTSLRCGS CIRKPLLCCK CAYDHVMSTD HKYVLSVSPY VCNSPGCDVN
AIBV        LQSCGVCVVC NSQTILRCGN CIRKPFLCCK CCYDHVMHTD HKNVLSINPY ICSQLGCGEA
SARS CoV    LQAVGACVLC NSQTSLRCGA CIRRPFLCCK CCYDHVISTS HKLVLSVNPY VCNAPGCDVT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5585       5595       5605       5615       5625       5635
EMCR        DVKKLYLGGL NYYCTNHKPQ LSFPLCSAGN IFGLYKNSAT GSLDVEVFNR LATSDWTDVR
229E        DVTKLYLGGL NYYCVDHKPH LSFPLCSAGN VFGLYKSSAL GSMDIDVFNK LSTSDWSDIR
PEDV        DVTKLYLGGL SYWCHEHKPR LAFPLCSAGN VFGLYKNSAT GSPDVEDFNR IATSDWTDVS
TGEV        DVTKLFLGGL SYYCMNHKPQ LSFPLCANGN VFGLYKSSAV GSEAVEDFNK LAVSDWTNVE
OV43        DVTKLYLGGM SYYCEDHKPQ YSFKLVMNGL VFGLYKQSCT GSPYIDDFNR IASCKWTDVD
BoCoV       DVTKLYLGGM SYYCEDHKPQ YSFKLVMNGM VFGLYKQSCT GSPYIDDFNR IASCKWTDVD
MHV         DVTKLYLGGM SYYCEDHKPQ YSFKLVMNGM VFGLYKQSCT GSPYIEDFNK IASCKWTEVD
AIBV        DVTKLYLGGM SYFCGNHKPK LSIPLVSNGT VFGIYRANCA GSENVDDFNQ LATTNWSIVE
SARS CoV    DVTQLYLGGM SYYCKSHKPP ISFPLCANGQ VFGLYKNTCV GSDNVTDFNA IATCDWTNAG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5645       5655       5665       5675       5685       5695
EMCR        DYKLANDVKD TLRLFAAETI KAKEESVKSS YAFATLKEVV GPKELLLSWE SGKVKPPLNR
229E        DYKLANDAKE SLRLFAAETV KAKEESVKSS YAYATLKEIV GPKELLLLWE SGKAKPPLNR
PEDV        DYRLANDVKD SLRLFAAETI KAKEESVKSS YACATLHEVV GPKELLLKWE VGRPKPPLNR
TGEV        DYKLANNVKE SLKIFAAETV KAKEESVKSE YAYAVLKEVI GPKEIVLQWE ASKTKPPLNR
OV43        DYILANECTE RLKLFAAETQ KATEEAFKQS YASATIQEIV SERELILSWE IGKVKPPLNK
BoCoV       DYILANECTE RLKLFAAETQ KATEEAFKQS YASATIQEIV SERELILSWE IGKVKPPLNK
MHV         DYVLANECTE RLKLFAAETQ KATEESFKQC YASATIREIV SDRELILSWE IGKVRPPLNK
AIBV        PYILANRCSD SLRRFAAETV KATEELHKQQ FASAEVREVF SDRELILSWE PGKTRPPLNR
SARS CoV    DYILANTCTE RLKLFAAETL KATEETFKLS YGIATVREVL SDRELHLSWE VGKPRPPLNR

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5705       5715       5725       5735       5745       5755
EMCR        NSVFTCFQIS KDSKFQIGEF IFEKVEYGSD TVTYKSTVTT KLVPGMIFVL TSHNVQPLRA
229E        NSVFTCFQIT KDSKFQVGEF VFEKVDYGSD TVTYKSTATT KLVPGMLFIL TSHNVAPLRA
PEDV        NSVFTCYHIT KNTKFQIGEF VFEKAEYDND AVTYKTTATT KLVPGMVFVL TSHNVQPLRA
TGEV        NSVFTCFQIS KDTKIQLGEF VFEQSEYGSD SVYYKSTSTY KLTPGMIFVL TSHNVSPLKA
```

```
OV43        NYVFTGYHFT KNGKTVLGEY VFDKSELT-N GVYYRATTTY KLSVGDVFVL TSHSVANLSA
BoCoV       NYVFTGYHFT KNGKTVLGEY VFDKSELT-N GVYYRATTTY KLSVGDVFVL TSHSVANLSA
MHV         NYVFTGYHFT SNGKTVLGEY VFDKSELT-N GVYYRATTTY KLSVGDVFVL TSHSVANLSA
AIBV        NYVFTGYHFT RTSKVQLGDF TFEKGEGK-D VVYYKATSTA KLSVGDIFVL TSHNVVSLVA
SARS CoV    NYVFTGYRVT KNSKVQIGEY TFEKGDYG-D AVVYRGTTTY KLNVGDYFVL TSHTVMPLSA

                    5765        5775        5785        5795        5805        5815
EMCR        PTIANQEKYS SIYKLHPAFN VSDAYANLVP YYQLIGKQKI TTIQGPPGSG KSHCSIGLGL
229E        PTMANQEKYS TIYKLHPSFN VSDAYANLVP YYQLIGKQRI TTIQGPPGSG KSHCSIGIGV
PEDV        PTIANQERYS TIHKLHPAFN IPEAYSSLVP YYQLIGKQKI TTIQGPPGSG KSHCVIGLGL
TGEV        PILVNQEKYN TISKLYPVFN IAEAYNTLVP YYQMIGKQKF TTIQGPPGSG KSHCVIGLGL
OV43        PTLVPQENYS SIR-FASVYS VLETFQNNVV NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV
BoCoV       PTLVPQENYS SIR-FASVYS VLETFQNNVV NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV
MHV         PTLVPQENYT SIR-FASVYS VPETFQNNVP NYQHIGMKRY CTVQGPPGTG KSHLAIGLAV
AIBV        PTLCPQQTFS RFVNLRPNVM VPECFVNNIP LYHLVGKQKR TTVQGPPGSG KSHFAIGLAV
SARS CoV    PTLVPQEHYV RITGLYPTLN ISDEFSSNVA NYQKVGMQKY STLQGPPGTG KSHFAIGLAL

                    5825        5835        5845        5855        5865        5875
EMCR        YYPGARIVFV ACAHAAVDSL CAKAMTVYSI DKCTRIIPAR ARVECYSGFK PNNTSAQYIF
229E        YYPGARIVFT ACSHAAVDSL CAKAVTAYSV DKCTRIIPAR ARVECYSGFK PNNNSAQYVF
PEDV        YYPGARIVFT ACSHAAVDSL CVKASTAYSN DKCSRIIPQR ARVECYDGFK SNNTSAQYLF
TGEV        YYPQARIVYT ACSHAAVDAL CEKAAKNFNV DRCSRIIPQR IRVDCYTGFK PNNTNAQYLF
OV43        FYCTARVVYT AASHAAVDAL CEKAYKFLNI NDCTRIVPAK VRVECYDKFK INDTTRKYVF
BoCoV       YYCTARVVYT AASHAAVDAL CEKAYKFLNI NDCTRIVPAK VRVECYDKFK INDTTRKYVF
MHV         YYCTARVVYT AASHAAVDAL CEKAYKFLNI NDCTRIVPAK VRVDCYDKFK VNDTTRKYVF
AIBV        YFSSARVVFT ACSHAAVDAL CEKAFKFLKV DDCTRIVPQR TTVDCFSKFK ANDTGKKYIF
SARS CoV    YYPSARIVYT ACSHAAVDAL CEKALKYLPI DKCSRIIPAR ARVECFDKFK VNSTLEQYVF

                    5885        5895        5905        5915        5925        5935
EMCR        STVNALPECN ADIVVVDEVS MCTNYDLSVI NQRLSYKHIV YVGDPQQLPA PRVMITKGVM
229E        STVNALPEVN ADIVVVDEVS MCTNYDLSVI NQRISYKHIV YVGDPQQLPA PRVLISKGVM
PEDV        STVNALPECN ADIVVVDEVS MCTNYDLSVI NQRISYRHVV YVGDPQQLPA PRVMISRGTL
TGEV        CTVNALPEAS CDIVVVDEVS MCTNYDLSVI NSRLSYKHIV YVGDPQQLPA PRTLINKGVL
OV43        TTINALPEMV TDIVVVDEVS MLTNYELSVI NARIRAKHYV YIGDPAQLPA PRVLLSKGTL
BoCoV       TTINALPEMV TDIVVVDEVS MLTNYELSVI NARIRAKHYV YIGDPAQLPA PRVLLSKGTL
MHV         TTINALPELV TDIIVVDEVS MLTNYELSVI NSRVRAKHYV YIGDPAQLPA PRVLLNKGTL
AIBV        STINALPEVS CDILLVDEVS MLTNYELSFI NGKINYQYVV YVGDPAQLPA PRTLLN-GSL
SARS CoV    CTVNALPETT ADIVVFDEIS MATNYDLSVV NARLRAKHYV YIGDPAQLPA PRTLLTKGTL

                    5945        5955        5965        5975        5985        5995
EMCR        EPVDYNVVTQ RMCAIGPDVF LHKCYRCPAE IVNTVSELVY ENKFVPVKPA SKQCFKIFFK
229E        EPIDYNVVTQ RMCAIGPDVF LHKCYRCPAE IVNTVSELVY ENKFVPVKEA SKQCFKIFER
PEDV        EPKDYNVVTQ RMCALKPDVF LHKCYRCPAE IVRTVSEMVY ENQFIPVHPD SKQCFKIFCK
TGEV        QPQDYNVVTK RMCTLGPDVF LHKCYRCPAE IVKTVSALVY ENKFVPVNPE SKQCFKMFVK
OV43        EPKYFNTVTK LMCCLGPDIF LGTCYRCPKE IVDTVSALVY ENKLKAKNES SSLCFKVYYK
BoCoV       EPKYFNTVTK LMCCLGPDIF LGTCYRCPKE IVDTVSALVY ENKLKAKNES SSLCFKVYYK
MHV         EPRYFNSVTK LMCCLGPDIF LGTCYRCPKE IVDTVSALVY HNKLKAKNDN SSMCFKVYYK
AIBV        SPKDYNVVTN LMVCVKPDIF LAKCYRCPKE IVDTVSTLVY DGKFIANNPE SRECFKVIVN
SARS CoV    EPEYFNSVCR LMKTIGPDMF LGTCRRCPAE IVDTVSALVY DNKLKAHKDK SAQCFKMFYK

                    6005        6015        6025        6035        6045        6055
EMCR        ---GNVQVDN GSSINRKQLE IVKLFLVKNP SWSKAVFISP YNSQNYVASR FLGLQIQTVD
229E        ---GSVQVDN GSSINRRQLD VVKRFIHKNS TWSKAVFISP YNSQNYVAAR LLGLQTQTVD
PEDV        ---GNVQVDN GSSINRRQLD VVRMFLAKNP RWSKAVFISP YNSQNYVASR LLGLQIQTVD
TGEV        ---GQVQIES NSSINNKQLE VVKAFLAHNP KWRKAVFISP YNSQNYVARR LLGLQTQTVD
OV43        ---GVTTHES SSAVNMQQIY LINKFLKANP LWHKAVFISP YNSQNFAAKR VLGLQTQTVD
BoCoV       ---GVTTHES SSAVNMQQIY LINKFLKANP LWHKAVFISP YNSQNFAAKR VLGLQTQTVD
MHV         ---GQTTHES SSAVNMQQIY LISKFLKANP SWSNAVFISP YNSQNYVAKR VLGLQTQTVD
AIBV        NGNSDVGHES GSAYNTTQLE FVKDFVCRNK QWREAIFISP YNAMNQRAYR MLGLNVQTVD
SARS CoV    ---GVITHDV SSAINRPQIG VVREFLTRNP AWRKAVFISP YNSQNAVASK ILGLPTQTVD

                    6065        6075        6085        6095        6105        6115
EMCR        SSQGSEYDYV IYAQTSDTAH ACNVNRFNVA ITRAKKGIFC VMCDKT-LFD SLKFFEIKHA
229E        SAQGSEYDYV IFAQTSDTAH ACNANRFNVA ITRAKKGIFC IMSDRT-LFD ALKFFEITMT
PEDV        SSQGSEYDYV IYAQTSDTAH ASNVNRFNVA ITRAKKGILC IMCDRS-LFD LLKFFELKLS
TGEV        SAQGSEYDYV IYTQTSDTQH ATNVNRFNVA ITRAKVGILC IMCDRT-MYE NLDFYELKDS
OV43        SAQGSEYDYV IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSNMQ-LFE ALQFTTLTLD
BoCoV       SAQGSEYDYV IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSNMQ-LFE ALQFTTLTLD
MHV         SAQGSEYDFV IYSQTAETAH SVNVNRFNVA ITRAKKGILC VMSNMQ-LFE ALQFTTLTVD
AIBV        SSQGSEYDYV IFCVTADSQH ALNINRFNVA LTRAKRGILV VMRQRDELYS ALKFTELDSE
SARS CoV    SSQGSEYDYV IFTQTTETAH SCNVNRFNVA ITRAKIGILC IMSDRD-LYD KLQFTSLEIP

                    6125        6135        6145        6155        6165        6175
```

```
EMCR       -----DLHSS -QVCGLFKNC TRTPLNLPPT HAHTFLSLSD QFKTTGDLAV QIGS-N-NVC
229E       -----DLQSE -SSCGLFKDC ARNPIDLPPS HATTYLSLSD RFKTSGDLAV QIGN-N-NVC
PEDV       -----DLQAN -EGCGLFKDC SRGDDLLPPS HANTFMSLAD NFKTDQYLAV QIGV-N-GPI
TGEV       KI---GLQAK PETCGLFKDC SKSEQYIPPA YATTYMSLSD NFKTSDGLAV NIG--T-KDV
OV43       KVPQAVETKV QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV
BoCoV      KVPQAVETRV QCSTNLFKDC SKSYSGYHPA HAPSFLAVDD KYKATGDLAV CLGIGD-SAV
MHV        KIN---NPRL QCTTNLFKDC SRSYAGYHPA HAPSFLAVDD KYKVGGDLAV CLNVAD-SAV
AIBV       T------S-- LQGTGLFKIC NKEFSGVHPA YAVTTKALAA TYKVNDELAA LVNVEAGSEI
SARS CoV   RRN-VATLQA ENVTGLFKDC SKIITGLHPT QAPTHLSVDI KFKTEG-LCV DIPGIP-KDM

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6185       6195       6205       6215       6225       6235
EMCR       TYEHVISFMG FRFDISIPGS HSLFCTRDFA IRNVRGWLGM DVESAHVCGD NIGTNVPLQV
229E       TYEHVISYMG FRFDVSMPGS HSLFCTRDFA MRHVRGWLGM DVEGAHVTGD NVGTNVPLQV
PEDV       KYEHVISFMG FRFDINIPNH HTLFCTRDFA MRNVRGWLGF DVEGAHVVGS NVGTNVPLQL
TGEV       KYANVISYMG FRFEANIPGY HTLFCTRDFA MRNVRAWLGF DVEGAHVCGD NVGTNVPLQL
OV43       TYSRLISLMG FKLDVTLDGY CKLFITKEEA VKRVRAWVGF DAEGAHATRD SIGTNFPLQL
BoCoV      TYSRLISLMG FKLDVTLDGY CKLFITKEEA VKRVRAWVGF DAEGAHATRD SIGTNFPLQL
MHV        TYSRLISLMG FKLDLTLDGY CKLFITRDEA IRRVRAWVGF DAEGAHATRD SIGTNFPLQL
AIBV       TYKHLISLLG FKMSVNVEGC HNMFITRDEA IRNVRGWVGF DVEATHACGT NIGTNLPFQV
SARS CoV   TYRRLISMMG FKMNYQVNGY PNMFITREEA IRHVRAWIGF DVEGCHATRD AVGTNLPLQL

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6245       6255       6265       6275       6285       6295
EMCR       GFSNGVNFVV QTEGCVSTNF GDVIKPVCAK SPPGEQFRHL VPFLRKGQPW LIVRRRIVQM
229E       GFSNGVDFVA QPEGCVLTNT GSVVKPVRAR APPGEQFTHI VPLLRKGQPW SVLRKRIVQM
PEDV       GFSNGVDFVV RPEGCVVTES GDYIKPVRAR APPGEQFAHL LPLLKRGQPW DVVRKRIVQM
TGEV       GFSNGVDFVV QTEGCVITEK GNSIEVVKAR APPGEQFAHL IPLMRKGQPW HIVRRRIVQM
OV43       GFSTGIDFVV EATGLFADRD GYSFKKAVAK APPGEQFKHL IPLMTRGHRW DVVRPRIVQM
BoCoV      GFSTGIDFVV EATGLFADRD GYSFKKAVAK APPGEQFKHL IPLMTRGQRW DVVRPRIVQM
MHV        GFSTGIDFVV EATGMFAERD GYVFKKAVAR APPGEQFKHL VPLMSRGQKW DVVRIRIVQM
AIBV       GFSTGADFVV TPEGLVDTSI GNNFEPVNSK APPGEQFNHL RVLFKSAKPW HVIRPRIVQM
SARS CoV   GFSTGVNLVA VPTGYVDTEN NTEFTRVNAK PPPGDQFKHL IPLMYKGLPW NVVRIKIVQM

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6305       6315       6325       6335       6345       6355
EMCR       ISDYLSNLSD ILVFVLWAGS LELTTMRYFV KIGP-IKYCY CGNSATCYNS VSNEYCCFKH
229E       IADFLAGSSD VLVFVLWAGG LELTTMRYFV KIGA-VKHCQ CGTVATCYNS VSNDYCCFKH
PEDV       CSDYLANLSD ILIFVLWAGG LELTTMRYFV KIGP-SKSCD CGKVATCYNS ALHTYCCFKH
TGEV       VCDYFDGLSD ILIFVLWAGG LELTTMRYFV KIGR-PQKCE CGKSATCYSS SQSVYACFKH
OV43       FADHLIDLSD CVVLVTWAAN FELTCLRYFA KVGREISCNV CTKRATVYNS RTGYYGCWRH
BoCoV      FADHLIDLSD CVVLVTWAAN FELTCLRYFA KVGREISCNV STKRATAYNS RTGYYGCWRH
MHV        LSDHLVDLAD SVVLVTWAAS FELTCLRYFA KVGKEVVCSV CNKRATCFNS RTGYYGCWRH
AIBV       LADNLCNVSD CVVFVTWCHG LELTTLRYFV KIGK-EQVCS CGSRATTFNS HTQAYACWKH
SARS CoV   LSDTLKGLSD RVVFVLWAHG FELTSMKYFV KIGPERTCCL CDKRATCFST SSDTYACWNH

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6365       6375       6385       6395       6405       6415
EMCR       ALGCDYVYNP YAFDIQQWGY VGSLSQNHHT FCNIHRNEHD ASGDAVMTRC LAVHDCFVKN
229E       ALGCDYVYNP YVIDIQQWGY VGSLSTNHHA ICNVHRNEHV ASGDAIMTRC LAVYDCFVKN
PEDV       ALGCDYLYNP YCIDIQQWGY KGSLSLNHHE HCNVHRNEHV ASGDAIMTRC LAIHDCFVKN
TGEV       ALGCDYLYNP YCIDIQQWGY TGSLSMNHHE VCNIHRNEHV ASGDAIMTRC LAIHDCFVKR
OV43       SVTCDYLYNP LIVDIQQWGY IGSLSSNHDL YCSVHKGAHV ASSDAIMTRC LAVYDCFCNN
BoCoV      SVTCDYLYNP LIVDIQQWGY IGSLSSNHDL YCSVHKGAHV ASSDAIMTRC LAVYDCFCNN
MHV        SYSCDYLYNP LIVDIQQWGY TGSLTSNHDL ICSVHKGAHV ASSDAIMTRC LAVHDCFCKS
AIBV       CLGFDFVYNP LLVDIQQWGY SGNLQFNHDL HCNVHGHAHV ASVDAIMTRC LAINNAFCQD
SARS CoV   SVGFDYVYNP FMIDVQQWGF TGNLQSNHDQ HCQVHGNAHV ASCDAIMTRC LAVHECFVKR

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6425       6435       6445       6455       6465       6475
EMCR       VDWTVTYPFI ANEKFINGCG RNVQGHVVRA ALKLYKPSVI HDIGNPKGVR CA-VTDAKWY
229E       VDWSITYPMI ANENAINKGG RTVQSHIMRA AIKLYNPKAI HDIGNPKGIR CA-VTDAKWY
PEDV       VDWSITYPFI GNEAVINKSG RIVQSHTMRS VLKLYNPKAI YDIGNPKGIR CA-VTDAKWF
TGEV       VDWSIVYPFI DNEEKINKAG RIVQSHVMKA ALKIFNPAAI HDVGNPKGIR CA-TTPIPWF
OV43       INWNVEYPII SNELSINTSC RVLQRVILKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK
BoCoV      INWNVEYPII SNELSINTSC RVLQRVMLKA AMLCNRYTLC YDIGNPKAIA CV--KDFDFK
MHV        VNWSLEYPII SNEVSVNTSC RLLQRVMFRA AMLCNRYDVC YDIGNPKGLA CV--KGYDFK
AIBV       VNWDLTYPHI ANEDEVNSSC RYLQRMYLNA CVDALKVNVV YDIGNPKGIK CVRRGDVNFR
SARS CoV   VDWSVEYPII GDELRVNSAC RKVQHMVVKS ALLADKFPVL HDIGNPKAIK CVPQAEVEWK

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              6485       6495       6505       6515       6525       6535
EMCR       CYDKQPVNSN ---VKLLDYD YATHG--QLD GLCLFWNCNV DMYPEFSIVC RFDTRTRSVF
229E       CYDKNPINSN ---VKTLEYD YMTHG--QMD GLCLFWNCNV DMYPEFSIVC RFDTRTRSTL
PEDV       CFDKNPTNSN ---VKTLEYD YITHG--QFD GLCLFWNCNV DMYPEFSVVC RFDTRCRSPL
TGEV       CYDRDPINNN ---VRCLDYD YMVHG--QMN GLMLFWNCNV DMYPEFSIVC RFDTRTRSKL
OV43       FYDAQPIVKS ---VKTLLYS FEAHKDSFKD GLCMFWNCNV DKYPPNAVVC RFDTRVLNNL
BoCoV      FYDAQPIVKS ---VKTLLYF FEAHKDSFKD GLCMFWNCNV DKYPPNAVVC RFDTRVLNNL
MHV        FYDASPVVKS ---VKQFVYK YEAHKDQFLD GLCMFWNCNV DKYPANAVVC RFDTRVLNKL
AIBV       FYDKNPIVRN ---VKQFEYD YNQHKDKFAD GLCMFWNCNV DCYPDNSLVC RYDTRNLSVF
```

```
                SARS CoV    FYDAQPCSDK AYKIEELFYS YATHHDKFTD GVCLFWNCNV DRYPANAIVC RFDTRVLSNL

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6545       6555       6565       6575       6585       6595
     5          EMCR        NLEGVNGGSL YVNKHAFHTP AYDKRAFVKL KPMPFFYFDD SDCDVVQ--- -EQVNYVPLR
                229E        NLEGVNGGSL YVNNHAFHTP AYDKRAMAKL KPAPFFYYDD GSCEVVH--- -DQVNYVPLR
                PEDV        NLEGCNGGSL YVNNHAFHTP AFDKRAFAKL KPMPFFFYDD TECDKLQ--- -DSINYVPLR
                TGEV        SLEGCNGGAL YVNNHAFHTP AYDRRAFAKL KPMPFFYYDD SNCELVD--- -GQPNYVPLK
                OV43        NLPGCNGGSL YVNKHAFHTK PFARAAFEHL KPMPFFYYSD TPCVYMDGMD AKQVDYVPLK
    10          BoCoV       NLPGCNGGSL YVNKHAFHTK PFSRAAFEHL KPMPFFYYSD TPCVYMDGMD AKQVDYVPLK
                MHV         NLPGCNGGSL YVNKHAFHTS PFTRAAFENL KPMPFFYYSD TPCVYMEGME SKQVDYVPLR
                AIBV        NLPGCNGGSL YVNKHAFYTP KFDRISFRNL KAMPFFFYDS SPCETIQVDG -VAQDLVSLA
                SARS CoV    NLPGCDGGSL YVNKHAFHTP AFDKSAFTNL KQLPFFYYSD SPCESHGKQV VSDIDYVPLK

    15                      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6605       6615       6625       6635       6645       6655
                EMCR        ASSCVTRCNI GGAVCSKHAN LYQKYVEAYN TFTQAGFNIW VPHSFDVYNL WQIFIET-NL
                229E        ATNCITKCNI GGAVCSKHAN LYRAYVESYN IFTQAGFNIW VPTTFDCYNL WQTFTEV-NL
                PEDV        ASNCITKCNV GGAVCSKHCA MYHSYVNAYN TFTSAGFTIW VPTSFDTYNL WQTFSN--NL
    20          TGEV        SNVCITKCNI GGAVCKKHAA LYRAYVEDYN IFMQAGFTIW CPQNFDTYML WHGFVNSKAL
                OV43        SATCITRCNL GGAVCLKHAE EYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTK---L
                BoCoV       SATCITRCNL GGAVCLKHAE EYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTK---L
                MHV         SATCITRCNL GGAVCLKHAE DYREYLESYN TATTAGFTFW VYKTFDFYNL WNTFTR---L
                AIBV        TKDCITKCNI GGAVCKKHAQ MYAEFVTSYN AAVTAGFTFW VTNKLNPYNL WKSFSA---L
    25          SARS CoV    SATCITRCNL GGAVCRHHAN EYRQYLDAYN MMISAGFSLW IYKQFDTYNL WNTFTR---L

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6665       6675       6685       6695       6705       6715
                EMCR        QSLENIAFNV VKKGCFTGVD GELPVAVVND KVFVRYGDVD NLVFTNKTTL PTNVAFELFA
    30          229E        QGLENIAFNV VNKGSFVGAD GELPVAISGD KVFVRDGNTD NLVFVNKTSL PTNIAFELFA
                PEDV        QGLENIAFNV LKKGSFVGDE GELPVAVVND KVLVRDGTVD TLVFTNKTSL PTNVAFELYA
                TGEV        QSLENVAFNV VKKGAFTGLK GDLPTAVIAD KIMVRDGPTD KCIFTNKTSL PTNVAFELYA
                OV43        QSLENVVYNL VKTGHYTGQA GEMPCAIIND KVVAKIDKED VVIFINNTTY PTNVAVELFA
                BoCoV       QSLENVVYNL VKTGHYTGQA GEMPCAIIND KVVAKIDKED VVIFINNTTY PTNVAVELFA
    35          MHV         QSLENVVYNL VNAGHFDGRA GELPCAVIGE KVIAKIQNED VVVFKNNTPF PTNVAVELFA
                AIBV        QSIDNIAYNM YKGGHYDAIA GEMPTVITGD KVFVIDQGVE KAVFVNQTTL PTSVAFELYA
                SARS CoV    QSLENVAYNV VNKGHFDGHA GEAPVSIINN AVYTKVDGID VEIFENKTTL PVNVAFELWA

    40                      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6725       6735       6745       6755       6765       6775
                EMCR        KRKMGLTPPL SILKNLGVVA TYKFVLWDYE AERPFTSYTK SVCKYTDFN- --------EDV
                229E        KRKVGLTPPL SILKNLGVVA TYKFVLWDYE AERPLTSFTK SVCGYTDFA- --------EDV
                PEDV        KRKVGLTPPI TILRNLGVVC TSKCVIWDYE AERPLTTFTK DVCKYTDFE- --------GDV
                TGEV        KRKLGLTPPL TILRNLGVVA TYKFVLWDYE AERPFSNFTK QVCSYTDLD- --------SEV
    45          OV43        KRSVRHHPEL KLFRNLNIDV CWKHVIWDYA RESIFCSNTY GVCMYTDLK- ------FIDKL
                BoCoV       KRSIRHHPEL KLFRNLNIDV CWKHVIWDYA RESIFCSNTY GVCMYTDLK- ------LIDKL
                MHV         KRSIRPHPEL KLFRNLNIDV CWSHVLWDYA KDSVFCSSTY KVCKYTDLQ- ------CIESL
                AIBV        KRNIRTLPNN RILKGLGVDV TNGFVIWDYA NQTPLYRNTV KVCAYTDIE- -------PNGL
                SARS CoV    KRNIKPVPEI KILNNLGVDI AANTVIWDYK REAPAHVSTI GVCTMTDIAK KPTESACSSL
    50

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6785       6795       6805       6815       6825       6835
                EMCR        CVCFDNSIQG SYERFTLTTN AVLFSTVVIK N----LTPIK LNFGMLNGMP VSSIKSDKGV
                229E        CTCYDNSIQG SYERFTLSTN AVLFSATAVK TGGKSLPAIK LNFGMLNGNA IATVKSEDGN
    55          PEDV        CTLFDNSIVG SLERFSMTQN AVLMSLTAVK K----LTGIK LTYGYLNGVP VN---THED-
                TGEV        VTCFDNSIAG SFERFTTTRD AVLISNNAVK G----LSAIK LQYGLLNDLP VS---TVGN-
                OV43        NVLFDGRDNG ALEAFKRSNN GVYISTTKVK S--------LS MIRGPPRAEL NGVVVDKVGD
                BoCoV       NVLFDGRDNG ALEAFKRSNN GVYISTTKVK S--------LS MIRGPPRAEL NGVVVDKVGD
                MHV         NVLFDGRDNG ALEAFKKCRD GVYINTTKIK S--------LS MIKGPQRADL NGVVVEKVGD
    60          AIBV        VVLYDDR-YG DYQSFLAADN AVLVSTQCYK R--------YS YVEIPSNLLV QNGMPLKDG-
                SARS CoV    TVLFDGRVEG QVDLFRNARN GVLITEGSVK G--------LT PSKGPAQASV NGVTLIGES-

                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
    65            6845       6855       6865       6875       6885       6895
                EMCR        EKLVNWYTYV RKNGQFQDHY DG--------- ---------- ---------- ------FYTQ
                229E        IKNINWFVYV RKDGKPVDHY DG--------- ---------- ---------- ------FYTQ
                PEDV        -KPFTWYIYT RKNGKFEDYP DG--------- ---------- ---------- ------FYTQ
                TGEV        -KPVTWYIYV RKNGEYVEQI DS--------- ---------- ---------- ------YFTQ
                OV43        -TDCVFYFAV RKEGQDVIFS QFDSLGVSSN QSPQGNLGSN GKPGNVGGND ALSISTIFTQ
    70          BoCoV       -TDCVFYFAV RKEGQDVIFS QFDSLRVSSN QSPQGNLGSN -EPGNVGGND ALATSTIFTQ
                MHV         -SDVEFWFAM RRDGDDVIFS RTGSLEPSHY RSPQGNPGGN -RVGDLSGNE ALARGTIFTQ
                AIBV        ---ANLYVYK RVNGAFVTLP N--------- ---------- ---------- -----TINTQ
                SARS CoV    -VKTQFNYFK KVDG--IIQ- ---------- QLP------- ---------- ----ETYFTQ
    75
                            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                              6905       6915       6925       6935       6945       6955
                EMCR        GRNLSDFTPR SDMEYDFLNM DMGVFINKYG LEDFNFEHVV YGDVSKTTLG GLHLLISQFR
                229E        GRNLQDFLPR STMEEDFLNM DIGVFIQKYG LEDFNFEHVV YGDVSKTTLG GLHLLISQVR
                PEDV        GRTTADFSPR SDMEKDFLSM DMGLFINKYG LEDYGFEHVV YGDVSKTTLG GLHLLISQVR
    80          TGEV        GRTFETFKPR STMEEDFLSM DTTLFIQKYG LEDYGFEHVV FGDVSKTTIG GMHLLISQVR
```

```
     OV43        SRVISSFTCR TDMEKDFIAL DQDVFIQKYG LEDYAFEHIV YGNFNQKIIG GLHLLIGLYR
     BoCoV       SRVISSFTCR TDMEKDFIAL DQDVFIQKYG LEDYAFEHIV YGNFNQKIIG GLHLLIGLYR
     MHV         SRFLSSFAPR SEMEKDFMDL DEDVFIAKYS LQDYAFEHVV YGSFNQKIIG GLHLLIGLAR
     AIBV        GRSYETFEPR SDIERDFLAM SEESFVERYG -KDLGLQHIL YGEVDKPQLG GLHTVIGMYR
5    SARS CoV    SRDLEDFKPR SQMETDFLEL AMDEFIQRYK LEGYAFEHIV YGDFSHGQLG GLHLMIGLAK

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    6965       6975       6985       6995       7005       7015
     EMCR        LSKMGVLKAD DFVTASDTTL RCCTVTYLNE LSSKVVCTYM DLLLDDFVTI LK---SLDLG
10   229E        LSKMGILKAE EFVAASDITL KCCTVTYLND PSSKTVCTYM DLLLDDFVSV LK---SLDLT
     PEDV        LACMGVLKID EFVSSNDSTL KSCTVTYADN PSSKMVCTYM DLLLDDFVSI LK---SLDLS
     TGEV        LAKMGLFSVQ EFMNNSDSTL KSCCITYADD PSSKNVCTYM DILLDDFVTI IK---SLDLN
     OV43        RQQTSNLVVQ EFVS-YDSSI HSYFITDEKS GGSKSVCTVI DILLDDFVAL VK---SLNLN
     BoCoV       RQQTSNLVIQ EFVS-YDSSI HSYFITDEKS GGSKSVCTVI DILLDDFVAL VK---SLNLN
15   MHV         RQQKSNLVIQ EFVP-YDSSI HSYFITDENS GSSKSVCTVI DLLLDDFVDI VK---SLNLN
     AIBV        LLRANKLNAK SVTN-SDSDV MQNYFVLSDN GSYKQVCTVV DLLLDDFLEL LRNILKEYGT
     SARS CoV    RSQDSPLKLE DFIP-MDSTV KNYFITDAQT GSSKCVCSVI DLLLDDFVEI IK---SQDLS

20               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    7025       7035       7045       7055       7065       7075
     EMCR        VISKVHEVII DNKPYRWMLW CKDNHLSTFY PQLQS-AEWK CGYAMPQIYK LQRMCLEPCN
     229E        VVSKVHEVII DNKPWRWMLW CKDNAVATFY PQLQS-AEWK CGYSMPGIYK TQRMCLEPCN
     PEDV        VVSKVHEVMV DCKMWRWMLW CKDHKLQTFY PQLQA-SEWK CGYSMPSIYK IQRMCLEPCN
     TGEV        VVSKVVDVIV DCKAWRWMLW CENSHIKTFY PQLQS-AEWN PGYSMPTLYK IQRMCLERCN
25   OV43        CVSKVVNVNV DFKDFQFMLW CNDEKVMTFY PRLQAASDWK PGYSMPVLYK YLNSPMERVS
     BoCoV       CVSKVVNVNV DFKDFQFMLW CNDEKVMTFY PRLQAASDWK PGYSMPVLYK YLNSPMERVS
     MHV         CVSKVVNVNV DFKDFQFMLW CNEEKVMTFY PRLQAAADWK PGYVMPVLYK YLESPLERVN
     AIBV        NKSKVVTVSI DYHSINFMTW FEDGSIKTCY PQLQS--AWT CGYNMPELYK VQNCVMEPCN
     SARS CoV    VISKVVKVTI DYAEISFMLW CKDGHVETFY PKLQASQAWQ PGVAMPNLYK MQRMLLEKCD

30               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    7085       7095       7105       7115       7125       7135
     EMCR        LYNYGAGIKL PSGIMLNVVK YTQLCQYLNS TTMCVPHNMR VLHYGAGSDK GVAPGTTVLK
     229E        LYNYGAGLKL PSGIMFNVVK YTQLCQYFNS TTLCVPHNMR VLHLGAGSDY GVAPGTAVLK
35   PEDV        LYNYGAGVKL PDGIMFNVVK YTQLCQYLNS TTMCVPHHMR VLHLGAGSDK GVAPGTAVLR
     TGEV        LYNYGAQVKL PDGITTNVVK YTQLCQYLNT TTLCVPHKMR VLHLGAAGAS GVAPGSTVLR
     OV43        LWNYGKPVTL PTGCMMNVAK YTQLCQYLNT TTLAVPVNMR VLHLGAGSEK GVAPGSAVLR
     BoCoV       LWNYGKPVTL PTGCMMNVAK YTQLCQYLNT TTLAVPVNTR VLHLGAGSEK GVAPGSAVLR
     MHV         LWNYGKPITL PTGCLMNVAK YTQLCQYLNT TTLAVPANMR VLHLGAGSDK DVAPGSAVLR
40   AIBV        IPNYGVGITL PSGILMNVAK YTQLCQYLSK TTICVPHNMR VMHFGAGSDK GVAPGSTVLK
     SARS CoV    LQNYGENAVI PKGIMMNVAK YTQLCQYLNT LTLAVPYNMR VIHFGAGSDK GVAPGTAVLR

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    7145       7155       7165       7175       7185       7195
45   EMCR        RWLPPD---- ------AIII DNDINDYVSD ADFSITGDCA TVYLEDKFDL LISDMYDG--
     229E        RWLPHD---- ------AIVV DNDVVDYVSD ADFSVTGDCA TVYLEDKFDL LISDMYDG--
     PEDV        RWLPLD---- ------AIIV DNDSVDYVSD ADYSVTGDCS TLYLSDKFDL VISDMYDG--
     TGEV        RWLPDD---- ------AILV DNDLRDYVSD ADFSVTGDCT SLYIEDKFDL LVSDLYDG--
     OV43        QWLPAG---- ------TILV DNDLYPFVSD SVATYFGDCI TLPFDCQWDL IISDMYDP--
50   BoCoV       QWLPAGTILR QWLPAGTILV HNDLYPFVSD SVATYFGDCI TLPFDCQWDL IISDMYD---
     MHV         QWLPAG---- ------SILV DNDINPFVSD SVASYYGNCI TLPIACQWDL IISDMYDP--
     AIBV        QWLPEG---- ------TLLV DNDIVDYVSD AHVSVLSDCN KYNTEHKFDL VISDMYTDND
     SARS CoV    QWLPTG---- ------TLLV DSDLNDFVSD ADSTLIGDCA TVHTANKWDL IISDMYDP--

55               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    7205       7215       7225       7235       7245       7255
     EMCR        --RIKFCDGE NVSKDGFFTY LNGVIREKLA IGGSVAIKIT EYSWNKYLYE LIQRFAFWTL
     229E        --RTKAIDGE NVSKEGFFTY INGFICEKLA IGGSIAIKVT EYSWNKKLYE LVQRFSFWTM
     PEDV        --KIKSCDGE NVSKEGFFPY INGVITEKLA LGGTVAIKVT EFSWNKKLYE LIQKFEYWTM
60   TGEV        --STKSIDGE NTSKDGFFTY INGFIKEKLS LGGSVAIKIT EFSWNKDLYE LIQRFEYWTV
     OV43        --ITKNIGEY NVSKDGFFTY ICHMIRDKLA LGGSVAIKIT EFSWNAELYK LMGYFAFWTV
     BoCoV       --LLLDIGVH VVRCS---YI HCHMIRDKLA LGGSVAIKIT EFSWNAELYK LMGYFAFWTV
     MHV         --LTKNIGEY NVSKDGFFTY LCHLIRDKLA LGGSVAIKIT EFSWNAELYS LMGKFAFWTI
     AIBV        SKRKHEGVIA NNGNDDVFIY LSSFLRNNLA LGGSFAVKVT ETSWHEVLYD IAQDCAWWTM
65   SARS CoV    --RTKHVTKE NDSKEGFFTY LCGFIKQKLA LGGSIAVKIT EHSWNADLYK LMGHFSWWTA

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    7265       7275       7285       7295       7305       7315
     EMCR        FCTSVNTSSS EAFLIGINYL GDFIQGPFIA GNTVHANYIF WRNSTIMSLS YNSVLDLSKF
70   229E        FCTSVNTSSS EAFVVGINYL GDFAQGPFID GNIIHANYVF WRNSTVMSLS YNSVLDLSKF
     PEDV        FCTSVNTSSS EAFLIGVHYL GDFASGAVID GNTMHANYIF WRNSTIMTMS YNSVLDLSKF
     TGEV        FCTSVNTSSS EGFLIGINYL GPYCDKAIVD GNIMHANYIF WRNSTIMALS HNSVLDTPKF
     OV43        FCTNANASSS EGFLIGINYL CKPKV--EID GNVMHANYLF WRNSTVWNGG AYSLFDMAKF
     BoCoV       FCTNANASSS EGFLIGINYL GKPKV--EID GNVMHAIICF G---------- ----------
75   MHV         FCTNVNASSS EGFLIGINWL NRTRT--EID GKTMHANYLF WRNSTMWNGG AYSLFDMSKF
     AIBV        FCTAVNASSS EAFLIGVNYL GASEK-VKVS GKTLHANYIF WRNCNYLQTS AYSIFDVAKF
     SARS CoV    FVTNVNASSS EAFLIGANYL GKPKE--QID GYTMHANYIF WRNTNPIQLS SYSLFDMSKF

80               ....|....| ....|....| ....|....| ....|....| ....|....|
                    7325       7335       7345       7355       7365
```

```
EMCR        ECKHKATVVV TLKDSDVNDM VLSLIKSGRL LLRNSGRFGG FSNHLVSTK-
229E        NCKHKATVVV QLKDSDINEM VLSLVRSGKL LVRGNGKCLS FSNHLVSTK-
PEDV        NCKHKATVVV NLKDSSISDV VLGLLKNGKL LVRNNDAICG FSNHLVNVNK
TGEV        KCRCNNALIV NLKEKELNEM VIGLLRKGKL LIRNNGKLLN FGNHFVNTP-
OV43        PLKLAGTAVI NLRADQINDM VYSLLEKGKL LIRDTNKEVF VGDSLVNVI-
BoCoV       EIPQFGTGVL IACLIWLNSR LSWLVMP--- ---------- ----------
MHV         PLKVAGTAVV SLKPDQINDL VLSLIEKGKL LVRDTRKEVF VGDSLVNVK-
AIBV        DLRLKATPVV NLKTEQKTDL VFNLIKCGKL LVRDVGNTSF TSDSFVCTM-
SARS CoV    PLKLRGTAVM SLKENQINDM IYSLLEKGRL IIRENNRVVV SSDILVNN--
```

e. Putative Spike protein

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5         15         25         35         45         55
EMCR S      ----MKLFLI LLILP----L VSCFSTC--- ---------N SNASIS---- --------ML
229E S      ---------- ---------- ---------- ---------- ---------- ----------
PEDV        MRSLIYFWLL LPVLP----T LSLPQDV--- ---------T RCQSTTNF-- -----RRFFS
TGEV        ---MKKLFVV LVVMP----L IYGDNFP--- ---------C SKLTNRTIGN QWNLIETFLL
CaCoV       ---MIVLTLC LFLFL-YSSV SCTSNND--- ---------C VQVNVTQLPG NENIIKDFLF
FeCoV       ---MIVLVTC LLLLCSYHTV LSTTNNE--- ---------C IQVNVTQLAG NENLIRDFLF
Por Resp C  ---MKKLFVV LVVMP----L IYG------- ---------- ---------- ----------
OC43        -----MFLIL LISLPTAFAV IGDLKCTSDT SYINDKDTGP PPISTDTVDV TNGLGTYYVL
BoCoV       -----MFLIL LISLPMAFAV IGDLKCT--T VSINDVDTGA PSISTDIVDV TNGLGTYYVL
MHV         -----MLFVF LTLLPSSLGY IGDFRCIQ-L VNTDTSNASA PSVSTEVVDV SKGIGTYYVL
Rat CoV     -----MLFVF LTLLPSCLGY IGDFRCIN-L VNTRISNARA PSVSTEVVDV SKGLGTYYVL
PHEV        -----MFFIL LISLPSAFAV IGDLKCT--T SLINDVDTGV PSISSEVVDV TNGLGTFYVL
AIBV        ---------- ---------- ---------- ---------- ---------- ----------
SARS        ----MFIFLL FLTLTSG--- -SDLDR---- ---------- -CTTFDDVQA PNYTQHTSSM
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     65         75         85         95        105        115
EMCR S      QLG--VPDNS STIVTGLLP- ---VHWICAN QSTSSYPANG FFYIDVG-KH RSAFALHSGY
229E S      ---------- ---------- ---------- ---------- ---------- ----------
PEDV        KFN--VQAPA VVVLGGYLPS MNSSSWYCGT GIETASGVHG IFLSYIDSGQ GFEIGISQEP
TGEV        NYSSRLPPNS DVVLGDYFPT --VQPWFNCI RNDSNDLYVT LENLKALYWD YATENITWN-
CaCoV       QN---FKEEG SLVVGGYYP- --TEVWYNCS TTQQTTAYKY FSNIHAFYFD MEAMENSTGN
FeCoV       SN---FKEEG SVVVGGYYP- --TEVWYNCS RTARTTAFQY FNNIHAFYFV MEAMENSTGN
Por Resp C  ---------- ---------- ---------- ---------- --------D- ----------
OC43        DR---VYLNT TLFLNGYYPT SGSTYRNMAL KGSVLLSRLW FKPPFLSDFI NGIFAKVKNT
BoCoV       DR---VYLNT TLLLNGYYPT SGSTYRNMAL KGTLLLSRLW FKPPFLSDFI NGIFAKVKNT
MHV         DR---VYLNA TLLLTGYYPV DGSMYRNMAL TGINTISLNW YKPPFLSEFN DGIFAKVKNL
Rat CoV     DR---VYLNA TLLLTGYYPV DGSMYRNMAL MGTNTLSLNW FEPPFLSEFN DGIYAKVKNL
PHEV        DR---VYLNT TLLLNGYYPI SGATFRNMAL KGTRLLSTLW FKPPFLSPFN DGIFAKVKNS
AIBV        ---------- ---------- ---------- ---------- ---------- ----------
SARS        RG---VYYPD EIFRSDTLYL TQDLFLPFYS NVTGFHTINH TFGNPVIPFK DGIYFAATEK
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    125        135        145        155        165        175
EMCR S      YDANQYYIYL TNKIH----- ---------- LNAPVTLKIC KFGN------ ---TSFDFLS
229E S      ---------- ---------- ---------- ---------- ---------- ----------
PEDV        FDPSGYQLYL HKATNG---- ---------N TNAIARLRIC QFPDN----- --KTLGPTVN
TGEV        -HRQRLNVVV NGYPYSITV- -----TTTRN FNSAEGAIIC ICKGSPPTTT TESSLTCNWG
CaCoV       ARGKPLLVHV HGNPVSIIVY ISAYRDDVQF RPLLKHGLLC ITKN--DTVD YNSFTINQWR
FeCoV       ARGKPLLFHV HGEPVSVII- -SAYRDDVQQ RPLLKHGLVC ITKN--RHIN YEQFTSNQWN
Por Resp C  ---------- ---------- ---------- ---------- ---------- ----------
OC43        KVIKDRVMYS EFPAITIG-- -------STF VNTSYSVVVQ PRTINSTQDG YNKLQGLLEV
BoCoV       KVIKKGVMYS EFPAITIG-- -------STF VNTSYSVVVQ PHTTN----L DNKLQGLLEI
MHV         KASLPKDSIS YFPTIIIG-- -------SNF VTTSYTVVLE PYN------- -----GIIMA
Rat CoV     KASLPIGSAS YFPTIIIG-- -------SNF VNTSYTVVLE PYN------- -----GIIMA
PHEV        RFSKDGVIYS EFPAITIG-- -------STF VNTSYSIVVE PHTSL----I NGNLQGLLQI
AIBV        ---------- ---------- ---------- ---------- ---------- ----------
SARS        SNVVRGWVFG STMNNKSQS- -------VII INNSTNVVIR ACN----FEL CDNPFFAVSK
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    185        195        205        215        225        235
EMCR S      NVSTSHDCIV NLSFTEQL-- GVPLGITISG ETVRLHLYNA TRTFYVPAAY KLTKLSVKCY
229E S      ---------- ---------- ---------- ---------- --MFVLLVAY ALLHIAG---
PEDV        DVTTGRNCLF NKAIPAYMRD GKDIVVGITW DNDRVTVF-A DKIYHFYLKN DWSRVATRCY
TGEV        SECR-LNHKF PICPSNSEAN CGNMLYGLQW FADEVVAYLH GASYRISFEN QWSGTVTFGD
CaCoV       DICLGDDRKI PFSVVPTDN- -GTKLFGLEW NDDYVTAYIS DESHRLNINN NWFNNVTLLY
FeCoV       STCTGADRKI PFSVIPTDN- -GTKIYGLEW NDDFVTAYIS GRSYHLNINT NWFNNVTLLY
Por Resp C  --------KF ------P--- ---------- ---------- ---------- ----------
OC43        SVCQYNMCEY PQTICHPNLG --NHRKELWH LDTGVVSCLY KRNFTYDVNA DYLYFHFYQ-
BoCoV       SVCQYTMCEY PHTICHPKLG --NKRVELWH WDTGVVSCLY KRNFTYDVNA DYLYFHFYQ-
MHV         SICQYTICQL PYTDCKPNTG G-NKLIGFWH TELKSPVCIL KRNFTFNVNA EWLYFHFYQ-
```

```
                 SICQYTICQL PHTDCKPNTG G-NTLIGFWH TDLRPPVCIL KRNFTFNVNA EWLYFHFYQ-
Rat CoV          SVCQYTMCEY PHTICHPNLG --NQRIELWH YDTDVVSCLY RRNFTYDVNA DYLYFHFYQ-
PHEV             ---------- ---------- ---------- ---MLVTPLL LVTLLCALCS AVLYDSS---
AIBV             PMGTQTHTMI FDNAFNCTFE YISDAFSLDV SE-KSGNFKH LREFVFKNKD GFLYVYKG--
SARS

                    245        255        265        275        285        295
                 FSESCVF--- ----SVVNA- -T-ITVNVTT LNGRIVNYTV CDD---CNGY TDNIFSVQQD
EMCR S           ---------- ---------- -------CQT TNGLNTSYSV CNG---CVGY SENVFAVESG
229E S           NRRSCAM--- ----QYVYTP TY-YMLNVTS AGEDGIYYEP CTAN--CTGY AANVFATDSN
PEDV             MRATTLEVAG TLVDLWWFNP VYDVSYYRVN NKNGTTVVSN CTD--QCASY VANVFTTQPG
TGEV             SRTSTATWQH S--AAYVYQG VSNFTYYKLN KTAGLKSYEL CEDYEYCTGY ATNVFAPTSG
CaCoV            SRSSTATWEY S--AAYAYQG VSNFTYYKLN NTNGLKTYEL CEDYEHCTGY ATNVFAPTSG
FeCoV            ---------- ---------- ---------- ----TSVVSN CTD--QCASY VANVFTILPG
Por Resp C       EGG------- ---------- -TFYAYFTDT GVVTKFLFNV YLG---MALS HYYVMPLTCN
OC43             EGG------- ---------- -TFYAYFTDT GVVTKFLFNV YLG---TVLS HYYVLPLTCS
BoCoV            QGG------- ---------- -TFYAYYADV SSATTFLFSM YIG---DVLT QYFVLPYMCT
MHV              QGG------- ---------- -TFYAYYADV SSATTFLFSS YIG---AVLT QYFVLPYMCS
Rat CoV          EGG------- ---------- -TFYAYFTDT GFVTKFLFKL YLG---TVLS HYYVMPLTCN
PHEV             S--------- ---------- ---YVYYYQS AFRPPSGWHL QGG---AYAV VNISSEFNNA
AIBV             YQP------- ---------- IDVVRDLPSG FNTLKPIFKL PLG-INITNF RAILTAFSPA
SARS

                    305        315        325        335        345        355
                 GRIPNGFPFN NWFL-LTNGS TLVDGVSRLY QPLRLTCLWP VPGLKSSTGF VYFNATGSDV
EMCR S           GYIPSDFAFN NWFL-LTNTS SVVDGVVRSF QPLLLNCLWS VSGLRFTTGF VYFNGTG-RG
229E S           GHIPEGFSFN NWFL-LSNDS TLLHGKVVSN QPLLVNCLLA IPKIYGLGQF FSFNHTM-DG
PEDV             GFIPSDFSFN NWFL-LTNSS TLVSGKLVTK QPLLVNCLWP VPSFEEAAST FCFEGAG-FD
TGEV             GYIPDGFSFN NWFM-LTNSS TFVSGRFVTN QPLLVNCLWP VPSFGVAAQE FCFEGAQ-FS
CaCoV            GYIPDGFSFN NWFL-LTNSS TFVSGRFVTN QPLLINCLWP VPSFGVAAQE FCFEGAQ-FS
FeCoV            GFIPSDFSFN NWFL-LTNSS TLVNGKLVTK QPLLVNCLWP VPSFEEVAST FCFEGAD-FD
Por Resp C       SKVKNGFTLE YWVTPLTSRQ YLLAFNQDGI IFNAVDCMSD FMSEIKCKTQ SIAPPTG-VY
OC43             S----AMTLE YWVTPLTSKQ YLLAFNQDGV IFNAVDCKSD FMSEIKCKTL SIAPSTG-VY
BoCoV            LTTTGVFSPQ YWVTPLVKRQ YLFNFNQKGI ITSAVDCASS YTSEIKCKTQ SMNPNTG-VY
MHV              PTTSGVSSPQ YWVTPLVKRQ YLFNFNQKGI ITSAVDCASS YTSEIKCKTQ SMNPNTG-VY
Rat CoV          S----ALSLE YWVTPLTTRQ FLLAFDQDGV LYHAVDCASD FMSEIMCKTS SITPPTG-VY
PHEV             G------SSS GCTVGIIHGG RVVNASSIAM TAPSSGMAWS SSQFCTAHCN FSDTTVFVTH
AIBV             QDIWGTSAAA YFVGYLKPTT FMLKYDENGT ITDAVDCSQN PLAELKCSVK SFEIDKG-IY
SARS

                    365        375        385        395        405        415
                 NCNGYQHNSV ADVMRYNLNL SANSVDNLKS GVIVFKTLQY -DVLFYCSN- --SS-SGVLD
EMCR S           DCKGFSSDVL SDVIRYNLNF EEN----LRR GTILFKTSYG -VVVFYCTN- --NT-LVSGD
229E S           VCNGAAVDRA PEALRFNIND TSV---ILAE GSIVLHTALG TNLSFVCSN- --SSDPHLAI
PEDV             QCNGAVLNNT VDVIRFNLNF TTNVQSGKGA TVFSLNTTGG VTLEISCY-- TVSDSSFFSY
TGEV             QCNGVSLNNT VDVIRFNLNF TTDVQSGMGA TVFSLNTTGG VILEISCYND TVSESSFYSY
CaCoV            QCNGVSLNNT VDVIRFNLNF TADVQSGMGA TVFSLNTTGG VILEISCYSD TVSESSSYSY
FeCoV            QCNGAVLNNT VDVIRFNLNF TTNVQSGKGA TVFSLNTTGG VTLEISCYND TVSDSSFSSY
Por Resp C       ELNGYTVQPI ADVYRRKLNL PNCNIEAWLN DKSVPSPLNW ERKTFSNCNF NMSSLMSFIQ
OC43             ELNGYTVQPI ADVYRRIPNL PDCNIEAWLN DKSVPSPLNW ERKTFSNCNF NMSSLMSFIQ
BoCoV            DLSGYTVQPV GLVYRRVRNL PDCKIEEWLT AKSVPSPLNW ERKTFQNCNF DLSSLLRFVQ
MHV              DLSGYTVQPV GLVYRRVRNL PDCKIEEWLA ANTVPSPLNW ERKTFQNCNF NLSSLLRFVQ
Rat CoV          ELNGYTVQPV ATVYRRIPDL PNCDIEAWLN SKTVSSPLNW ERKIFSNCNF NMGRLMSFIQ
PHEV             CYKHGGCPLT GMLQQNLIRV SAMKNGQLFY NLTVSVAKYP TFRSFQCVN- --NLTSVYLN
AIBV             QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF
SARS

                    425        435        445        455        465        475
                 TTIPFGPSSQ PYYCFINSTI NTTHVSTFVG ILPPTVREIV VARTGQFYIN GFKYFDLGFI
EMCR S           AHIPFGTVLG NFYCFVNTTI GNETTSAFVG ALPKTVREFV ISRTGHFYIN GYRYFTLGNV
229E S           FAIPLGATEV PYYCFLKVDT YNSTVYKFLA VLPPTVREIV ITKYGDVYVN GFGYLHLGLL
PEDV             GEIPFGVTDG PRYCYV---H YNGTALKYLG TLPPSVKEIA ISKWGHFYIN GYNFFSTFPI
TGEV             GEIPFGVTDG PRYCYV---L YNGTALKYLG TLPPSVKEIA ISKWGHFYIN GYNFFSTFPI
CaCoV            GEIPFGITDG PRYCYV---L YNGTALKYLG TLPPSVKEIA ISKWGHFYIN GYNFFSTFPI
FeCoV            GEIPFGVTNG PRYCYV---L YNGTALKYLG TLPPSVKEIA ISKWGHFYIN GYNFFSTFPI
Por Resp C       ADSFTCNNID AAKIYG--MC FSSITIDKFA IPNGRKVDLQ LGNLGYLQSF NYRIDTTATS
OC43             ADSFTCNNID AAKIYG--MC FSSITIDKFA IPNGRKVDLQ LGNLGYLQSF NYRIDTTATS
BoCoV            AESLSCSNID ASKVYG--MC FGSISIDKFA IPNRRRVDLQ LGNSGFLQSF NYKIDTRATS
MHV              AESLSCSNID ASKVYG--MC FGSISIDKFA IPNSRRVDLQ LGKSGLLQSF NYKIDTRATS
Rat CoV          ADSFGCNNID ASRLYG--MC FGSITIDKFA IPNSRKVDLQ VGKSGYLQSF NYKIDTAVSS
PHEV             GDLVYTSNET IDVTSAG--V YFKAGGPITY KVMREVKALA YFVNGTAQDV ILCDGSPRGL
AIBV             FSTFKCYGVS ATKLND--LC FSNVYADSFV VKGDDVRQIA PGQTGVIADY NYKLPDDFMG
SARS

                    485        495        505        515        525        535
                 EAVNFNVT-- --TASATDFW TVAFATFVDV LVNVSATNIQ NLLYCDSPFE KLQCEHLQFG
EMCR S           EAVNFNVT-- --TAETTDFC TVALASYADV LVNVSQTSIA NIIYCNSVIN RLRCDQLSFD
229E S           DAVTINFTGH GTDDDVSGFW TIASTNFVDA LIEVQGTSIQ RILYCDDPVS QLKCSQVAFD
PEDV             DCISFNLT-- --TGDSDVFW TIAYTSYTEA LVQVENTAIT KVTYCNSHVN NIKCSQITAN
TGEV             DCIAFNLT-- --TGASGAFW TIAYTSYTEA LVQVENTAIK KVTYCNSHIN NIKCSQLTAN
CaCoV
```

```
FeCoV        GCISFNLT-- --TGVSGAFW TIAYTSYTEA LVQVENTAIK NVTYCNSHIN NIKCSQLTAN
Por Resp C   DCISFNLT-- --TGDSDVFW TIAYTSYTEA LVQVENTAIT NVTYCNSYVN NIKCSQLTAN
OC43         CQLYYNLP-- ---AANVSVS RFNPSTWNKR FGFIEDSVFK PRPAGVLTNH DVVYAQHCFK
BoCoV        CQLYYNLP-- ---AANVSVS RFNPSTWNRR FGFTEQFVFK PQPVGVFTHH DVVYAQHCFK
MHV          CQLYYSLA-- ---KNNVTVN NHNPSSWNRR YGFND----- -VATFGTGKH DVAYAEACFT
Rat CoV      CQLYYSLA-- ---QDNVTVI NHNPSSWNRR YGFND----- -VATFHSGEH DVAYAEACFT
PHEV         CQLYYSLP-- ---AANVSVT HYNPSSWNRR YGFNN----- -QSFGSRGLH DAVYSQQCFN
AIBV         LACQYNTG-- ---NFSDGFY PFTNSSLVKQ KFIVYR---- ----ENSVNT TCTLHNFIFH
SARS         CVLAWNTR-- -------NID ATSTGNYNYK YRYLR----- ---HG----- ----------

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                545        555        565        575        585        595
EMCR S       LQDGFY--SA --NFLDDNVL P-------ET YVALPIYYQH TDINFTATA- --SFGGSCYV
229E S       VPDGFY--ST --SPIQSVEL P-------VS IVSLPVYHKH TFIVLYVDFK PQSGGGKCFN
PEDV         LDDGFYPISS --RNLLSHEQ P-------IS FVTLPSFNDH SFVNITVSA- --AFGG-LSS
TGEV         LNNGFYPVSS --SEVG--LV N-------KS VVLLPSFYTH TIVNITIGLG -MKRSGYGQP
CaCoV        LQNGFYPVAS --SEVG--LV N-------KS VVLLPSFYSH TSVNITIDLG -MKRS-VTVT
FeCoV        LNNGFYPVAS --SEVG--FV N-------KS VVLLPSFFTY TAVNITIDLG -MKLSGYGQP
Por Resp C   LNNGFYPVSS --SEVG--SV N-------KS VVLLPSFLTH TIVNITIGLG -MKRSGYGQP
OC43         APKNFCPCKL NGS-CVGSGP G-----KNNG IGTCPAGTNY LTCDN----- ---------LC
BoCoV        APKNFCPCKL DGSLCVGNGP GIDAGYKNSG IGTCPAGTNY LTCHNAA--- ----QCDCLC
MHV          VGASYCPCAN P-SIVSPCTT G-----K-PN FANCPTGTSN RECTVMPLAN -NQFKCDCTC
Rat CoV      VGASYCPCAK P-STVYSCVT G-----K-PK SANCPTGTSN RECNVQASG- -FKSKCDCTC
PHEV         TPNTYCPCRT --SQCIG--- G-------AG TGTCPVGTTV RKCFAAVTK- --ATKCTCWC
AIBV         NETGANPNPS ---------- ---------G VQNIQTYQTK TAQSGYYNFN -------FSF
SARS         ---KLRPFER ---------- ---------D ISNVPFSPDG KPCTP----- ----------

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                605        615        625        635        645        655
EMCR S       CKPRQVNISL N----GNTSV CVRTSHFSIR YIYNRVKSGS P--------G DSSWHIYLKS
229E S       CYPAGVNITL ANFNETKGPL CVDTSHFTTK YVAVYANVG- ---------- --RWSASINT
PEDV         ANLVASDTTI N----GFSSF CVDTRQFTIT LFYNVTNSYG ---------- ---YVSKSQD
TGEV         IASTLSNITL PMQDHNTDVY CIRSDQFSVY VHSTCKSALW DNIFKRNCTD VLDATAVIKT
CaCoV        IASPLSNITL PMQDNNIDVY CIRSNQFSVY VHSTCKSSLW DNNFNSACTD VLDATAVIKT
FeCoV        IASTLSNITL PMQDNNTDVY CIRSNQFSVY VHSTCKSSLW DNIFNQDCTD VLEATAVIKT
Por Resp C   IASTLSNITL PMQDNNNDVY CVRSDQFSVY VHSTCKSVLW DNVFKRNCTD VLDATAVIKT
OC43         TPDPITFKAT GTYKCPQTKS LVGIGEHCSG LAVKSDYCGG ---------N SCTCRPQAFL
BoCoV        TPDPITSKST GPYKCPQTKY LVGIGEHCSG LAIKSDYCGG ---------N PCTCQPQAFL
MHV          NPSPLTTYDL --R-CLQARS MLGVGDHCEG LGVLEDKCGG S--------N TCNCSAHAFV
Rat CoV      NPSPLTTYDP --R-CLQARS MLGVGDHCEG LGILEDKCGG S--------N ICNCSADAFV
PHEV         QPDPSTYKGV NAWTCPQSKV SIQPGQHCPG LGLVEDDCSG ---------N PCTCKPQAFI
AIBV         LSSFVYKESN FMYGSYHPSC KFRLETINNG LWFNSLSVS- ---------- ---IAYGPLQ
SARS         ---------- -----P--AL NCYWPLNDY- ---------G ---------- ---FYTTTGI

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                665        675        685        695        705        715
EMCR S       GTCPFSFSKL NNFQKFKTIC FSTVEVPGSC NFPLEATW-- HYTSYTIVGA LYVTWSEGNS
229E S       GNCPFSFGKV NNFVKFGSVC FSLKDIPGGC AMPIVANW-- AYSKYYTIGS LYVSWSDGDG
PEDV         SNCPFTLQSV NDYLSFSKFC VSTSLLAGAC TIDLFGYP-- AFGSGVKLTS LYFQFTKGEL
TGEV         GTCPFSFDKL NNYLTFNKFC LSLSPVGANC KFDVAAR--- TRTNEQVVRS LYVIYEEGDN
CaCoV        GTCPFSFDKL NNYLTFNKFC LSLNPVGANC KLDVAAR--- TRTNEQVFGS LYVIYEEGDN
FeCoV        GTCPFSFDKL NNYLTFNKFC LSLSPVGANC KFDVAAR--- TRTNEQVVRS LYVIYEEGDN
Por Resp C   GTCPFSFDKL NNYLTFNKFC LSLSPVGANC KFDVAAR--- TRTNDQVVRS LYVIYEEGDS
OC43         GWSADSCLQG DKCNIFANFI LHDVNSGLTC STDLQKANTD IILGVCVNYD LYGILGQGIF
BoCoV        GWSVDSCLQG DRCNIFANFI FHDVNSGTTC STDLQKSNTD IILGVCVNYD LYGITGQGIF
MHV          GWAKDSCLAN GRCHIFSNLM LNGINSGTTC SMDLQLPNTE VVTGVCVKYD LYGITGQGIF
Rat CoV      GWAMDSCLSN ARCHIFSNLM LNGINSGTTC STDFQLPNTE VVTGVCVKYD LYGSTGQGVF
PHEV         GWSSETCLQN GRCNIFANFI LNDVNSGTTC STDLQQGNTI ITTDVCVNYD LYGITGQGIL
AIBV         GGCKQSVFKG RATCCYAYSY GGPSLCKGVY SGELDHN--- ---FECGLLV YVTKSGGSRI
SARS         GYQPYRVVVL S----FELLN APATVCGPKL STDLIKN--- ----QCVNFN FNGLTGTG-V

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                725        735        745        755        765        775
EMCR S       ITGVPYPVSG IREFSNLVLN NCTKYNIYDY VGTGIIRSSN QSLAGGITYV S---------
229E S       ITGVPQPVEG VSSFMNVTLD KCTKYNIYDV SGVGVIRVSN DTFLNGITYT S---------
PEDV         ITGTPKPLEG ITDVSFMTLD VCTKYTIYGF KGEGIITLTN SSILAGVYYT S---------
TGEV         IVGVPSDNSG VHDLSVLHLD SCTDYNIYGR TGVGIIRQTN RTLLSGLYYT S---------
CaCoV        IVGVPSDNSG LHDLSVLHLD SCTDYNIYGR TGVGIIRKTN STLLSGLYYT S---------
FeCoV        IVGVPSDNSG LHDLSVLHLD SCTDYNIYGR TGVGIIRRTN STLLSGLYYT S---------
Por Resp C   IVGVPSDNSG LHDLSVLHLD SCTDYNIYGR TGVGIIRQTN RTILSGLYYT S---------
OC43         VEVNATYYNS WQNLLYDSNG NLYGFRDYIT NRTFMIRSCY SGRVSAAFHA N---------
BoCoV        VEVNATYYNS WQNLLYDSNG NLYGFRDYLT NRTFMIRSCY SGRVSAAFHA N---------
MHV          KEVKADYYHS WQNLLYDVNG NLIGFRDFVA NKSYTIRSCY SGRVSAAYHQ D---------
Rat CoV      KEVKADYYNS WQNLLYDVNG NLNGFRDIVT NKTYLLRSCY SGRVSAAYHQ D---------
PHEV         IEVNATYYNS WQNLLYDSSG NLYGFRDYLS NRTFLIRSCY SGRVSAVFHA N---------
AIBV         QTATEPPVIT QNNYNNITLN TCVDYNIYGR TGQGFITNVT DSAVSYNYLA DAGLAILDTS
SARS         LTPSSKRFQP FQQFGRDVSD FTDSVRDPKT SEILDISPCS FGGVSVITPG TN-------A

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                785        795        805        815        825        835
```

```
EMCR S      NSGNLLGFKN VSTGNIFIVT PCNQPDQVAV YQQ-SIIGAM TAVNESRYGL QNLLQLPNFY
229E S      TSGNLLGFKD VTKGTIYSIT PCNPPDQLVV YQQ-AVVGAM LSENFTSYGF SNVVELPKFF
PEDV        DSGQLLAFKN VTSGAVYSVT PCSFSEQAAY VND-DIVGVI SSLSNS--TF NNTRELPGFF
TGEV        LSGDLLGFKN VSDGVIYSVT PCDVSAQAAV IDG-TIVGAI TSINSELLGL THWTTTPNFY
CaCoV       LSGDLLGFKN VSDGVVYSVT PCDVSAQAAV IDG-AIVGAM TSINSELLGL THWTTTPNFY
FeCoV       LSGDLLGFKN VSDGVIYSVT PCDVSAQAAV IDG-AIVGAM TSINSELLGL THWTTTPNFY
Por Resp C  LSGDLLGFTN VSDGVIYSVT PCDVSAQAAI IDG-TIVGAI TSINSELLGL THWTTTPNFY
OC43        SSEPALLFRN IKCNYVFNNS LTRQLQPINY FDS-YLGCVV NAYNSTAISV QTCDLTVGSG
BoCoV       SSEPALLFRN IKCNYVFNNT LSRQLQPINY FDS-YLGCVV NADNSTSSVV QTCDLTVGSG
MHV         APEPALLYRN LKCDYVFNNN ISREETPLNY FDS-YLGCVV NADNSTEEAV DACDLRMGSG
Rat CoV     APEPALLYRN LKCDYVFNNN ISREETPLNY FDS-YLGCVI NADNSTEQSV DACDLRMGSG
PHEV        SSEPALMFRN LKCSHVFNNT ILRQIQLVNY FDS-YLGCVV NAYNNTASAV STCDLTVGSG
AIBV        GSIDIFVVQG EYGLNYYKVN PCEDVNQQFV VSGGKLVGIL TSRNETGSQL LENQFYIKIT
SARS        SSEVAVLYQD VNCTDVSTAI HADQLTPAWR IYS-TGNNVF QTQAGCLIGA EHVDTSYECD

                  845       855       865       875       885       895
EMCR S      YVS-----NG GNN------- -----CTTAV MIYSNFGICA DGSLIPVRPR NSSDNGISAI
229E S      YAS-----NG TYN------- -----CTDAV LTYSSFGVCA DGSIIAVQPR NVSYDSVSAI
PEDV        YHS-----ND GSN------- -----CTEPV LVYSNIGVCK SGSIGYV-PS QYGQVKIAPT
TGEV        YYSI---YNY TNDRTRGTAI DSNDVDCEPV ITYSNIGVCK NGAFVFIN-V THSDGDVQPI
CaCoV       YYSI---YNY TNVMNRGTAI D-NDIDCEPI ITYSNIGVCK NGALVFIN-V THSDGDVQPI
FeCoV       YYSI---YNY TSERTRGTAI DSNDVDCEPV ITYSNIGVCK NGALVFIN-V THSDGDVQPI
Por Resp C  YYSI---YNY TNDKTRGTPI GSNDVDCEPV ITYSNIGVCK NGALVFIN-V THSDGDVQPI
OC43        YCVD---YSK NRR------- -----SRGAI TTGYRFTNFE PFTVNSVN-- ---DSLEPVG
BoCoV       YCVD---YST KRR------- -----SRRAI TTGYRFTNFE PFTVNSVN-- ---DSLEPVG
MHV         LCVN---YST SHR------- -----ARSSV STGYKLTTFE PFTVRIVN-- ---DSVESVD
Rat CoV     LCVN---YSI AHR------- -----ARRSV STGYKLTTFE PFTVSIVN-- ---DSVESVG
PHEV        YCVD---YVT ALR------- -----SRRSF TTGYRFTNFE PFAANLVN-- ---DSIEPVG
AIBV        NGTRRFRRSI TEN------- ----VANCPY VSYGKFCIKP DGSIATIVPK QLEQFVAPLF
SARS        IPIG---AGI CAS------- ----YHTVSL LRSTSQKSIV AYTMSLG--- ---ADSSIAY

                  905       915       925       935       945       955
EMCR S      -ITANLSIPS NWTTSVQVEY LQITSTPIVV DCATYVCNGN PRCKNLLKQY TSACKTIEDA
229E S      -VTANLSIPS NWTTSVQVEY LQITSTPIVV DCSTYVCNGN VRCVELLKQY TSACKTIEDA
PEDV        -VTGNISIPT NFSMSIRTEY LQLYNTPVSV DCATYVCNGN SRCKQLLTQY TAACKTIESA
TGEV        -STGNVTIPT NFTISVQVEY IQVYTTPVSI DCSRYVCNGN PRCNKLLTQY VSACQTIEQA
CaCoV       -STGNVTIPT NFTISVQVEY IQVYTTPVSI DCARYVCNGN PRCNKLLTQY VSACQTIEQA
FeCoV       -STGNVTIPT NFTISVQVEY MQVYTTPVSI DCARYVCNGN PRCNKLLTQY VSACQTIEQA
Por Resp C  -STGNVTIPT NFTISVQVEY IQVYTTPVSI DCSRYVCNGN PRCNKLLTQY VSACQTIEQA
OC43        -GLYEIQIPS EFTIGNMEEF IQTSSPKVTI DCAAFVCGDY AACKSQLVEY GSFCDNINAI
BoCoV       -GLYEIQIPS EFTIGNMEEF IQTSSPKVTI DCSAFVCGDY AACKSQLVEY GSFCDNINAI
MHV         -GLYELQIPT NFTIASHQEF VQTRSPKVTI DCAAFVCGGH TACRQQLVEY GSFCDNINAI --
Rat CoV     -GLYEMQIPT NFTIASHQEF IQTRSPKVTI DCAAFVCGDY TACRQQLVDY GSFCDNINAI
PHEV        -GLYEIQIPS EFTIGNLEEF IQTRSPKVTI DCATFVCGDY AACRQQLAEY GSFCENINAI
AIBV        NVTENVLIPN SFNLTVTDEY IQTRMDKVQI NCLQYVCGSS LDCRKLFQQY GPVCDNILSV
SARS        -SNNTIAIPT NFSISITTEV MPVSMAKTSV DCNMYICGDS TECANLLLQY GSFCTQLNRA

                  965       975       985       995      1005      1015
EMCR S      LRLSAHLETN DVSSMLTFDS NA-FSLANVT SFG------D YNLSSVLPQ- ----------
229E S      LRNSARLESA DVSEMLTFDK KA-FTLANVS SFG------D YNLSSVIPS- ----------
PEDV        LQLSARLESV EVNSMLTISE EA-LQLATIS SFNG----DG YNFTNVLGAS ----------
TGEV        LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----SS ETLDPIYKEW PNIGGSWLEG
CaCoV       LAMGARLENM EIDSMLFVSE NA-LKLASVE AFN-----ST ENLDPIYKEW PNIGGSWLGG
FeCoV       LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----ST ENLDPIYKEW PSIGGSWLGG
Por Resp C  LAMGARLENM EVDSMLFVSE NA-LKLASVE AFN-----SS ETLDPIYKEW PNIGGFWLEG
OC43        LTEVNELLDT TQLQVANSLM NG-VTLSTKL KDGVNFNVDD INFSPVLGCL G---------
BoCoV       LTEVNELLDT TQLQVANSLM NG-VTLSTKL KDGVNFNVDD INFSPVLGCL G---------
MHV         LGEVNNLIDT MQLQVASALI QG-VTLSSRL SDGIGGQIDD INFSPLLGCL G---------
Rat CoV     LGEVNNLIDT MQLQVASALI QG-VTLSSRL ADGISGQIDD INFSPLLGCL G---------
PHEV        LTEVNELLDT TQLQVANSLM NG-VTLSTKI KDGINFNVDD INFSPVLGCL G---------
AIBV        VNSVGQKEDM ELLNFYSSTK PAGFNTPVLS NVSTG----E FNISLLLTN- ----------
SARS        LSGIAAEQDR NTREVFAQVK QM-YKTPTLK YFG------G FNFSQILPDP ----------

                 1025      1035      1045      1055      1065      1075
EMCR S      ----RNIHSS RIAGRSALED LLFSKVVTSG LGTVDVDYKS CTKGLS--IA DLACAQYYNG
229E S      ----LPTSGS RVAGRSAIED ILFSKLVTSG LGTVDADYKK CTKGLS--IA DLACAQYYNG
PEDV        V--YDPASGR VVQKRSVIED LLFNKVVTNG LGTVDEDYKR CSNGRS--VA DLVCAQYYSG
TGEV        LKYILPSHNS KRKYRSAIED LLFDKVVTSG LGTVDEDYKR CTGGYD--IA DLVCAQYYNG
CaCoV       LKDILPSHNS KRKYRSAIED LLFDKVVTSG LGTVDEDYKR SAGGYD--IA DLVCARYYNG
FeCoV       LKDILPSHNS KRKYGSAIED LLFDKVVTSG LGTVDEDYKR CTGGYD--IA DLVCAQYYNG
Por Resp C  LKYILPSDNS KRKYRSAIED LLFSKVVTSG LGTVDEDYKR CTGGYD--IA DLVCAQYYNG
OC43        -SECSKASS- ----RSAIED LLFDKVKLSD VG-FVEAYNN CTGGAE--IR DLICVQSYKG
BoCoV       -SACNKVSS- ----RSAIED LLFSKVKLSD VG-FVEAYNN CTGGAE--IR DLICVQSYNG
MHV         -SDCGEVTMA AQTGRSAIED VLFDKVKLSD VG-FVEAYNN CTGGQE--VR DLLCVQSFNG
Rat CoV     -SDCSEGTKA AQ-GRSAIED VLFDKVKLSD VG-FVESYNN CTGGQE--VR DLLCVQSFNG
PHEV        -SECNRAST- ----RSAIED LLFDKVKLSD VG-FVQAYNN CTGGAE--IR DLICVQSYNG
```

```
AIBV        -----PSSRR K---RSLIED LLFTSVESVG LP-TNDAYKN CTAGPLGFFK DLACAREYNG
SARS        ----LKPTK- ----RSFIED LLFNKVTLAD AG-FMKQYGE CLGDIN--AR DLICAQKFNG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    1085       1095       1105       1115       1125       1135
EMCR S      IMVLPGVADA ERMAMYTGSL IGGMVLGGLT S----AAAIP FSLALQARLN YVALQTDVLQ
229E S      IMVLPGVADA ERMAMYTGSL IGGIALGGLT S----AVSIP FSLAIQARLN YVALQTDVLQ
PEDV        VMVLPGVVDA EKLHMYSASL IGGMALGGIT A----AAALP FSYAVQARLN YLALQTDVLQ
TGEV        IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAIP FAVAVQARLN YVALQTDVLN
CaCoV       IMVLPGVAND DKMTMYTASL TGGITLGALS GG---AVAIP FAVAVQARLN YVALQTDVLN
FeCoV       IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAIP FAVAVQARLN YVALQTDVLN
Por Resp C  IMVLPGVANA DKMTMYTASL AGGITLGALG GG---AVAIP FAVAVQARLN YVALQTDVLN
OC43        IKVLPPLLSE NQISGYTLAA TSASLFPLWT A----AAGVP FAVAVQARLN YVALQTDVLN
BoCoV       IKVLPPLLSV NQISGYTLAA TSASLFPPLS A----AVGVP FYLNVQYRIN GLGVTMDVLS
MHV         IKVLPPVLSE NQISGYTAGA TVSAMFP-WS A----AAGVP FYLNVQYRIN GIGVTMDVLS
Rat CoV     IKVLPPVLSE SQISGYTAGA TASAMFPPWS A----AAGVP FSLSVQYRIN GLGVTMNVLS
PHEV        IKVLPPLLSE NQISGYTLAA TAASLFPPWT A----AAGVP FALSVQYRIN GLGVTMNVLS
AIBV        LLVLPPIITA EMQALYTSSL VASMAFGGIT A----AGAIP FYLNVQYRIN GLGVTMDVLS
SARS        LTVLPPLLTD DMIAAYTAAL VSGTATAGWT FGAGAALQIP FAMQMAYRFN GIGVTQNVLY

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    1145       1155       1165       1175       1185       1195
EMCR S      ENQKILAASF NKAINNIVAS FSSVNDAITH TAEAIHTVTI ALNKIQDVVN QQGSALNHLT
229E S      ENQKILAASF NKAMTNIVDA FTGVNDAITQ TSQALQTVAT ALNKIQDVVN QQGNSLNHLT
PEDV        RNQQLLAESF NSAIGNITSA FESVKEAISQ TSKGLNTVAH ALTKVQEVVN SQGSALNQLT
TGEV        KNQQILASAF NQAIGNITQS FGKVNDAIHQ TSRGLATVAK ALAKVQDVVN IQGQALSHLT
CaCoV       KNQQILANAF NQAIGNITQA FGKVNDAIHQ TSKGLATVAK ALAKVQDVVN TQGQALSHLT
FeCoV       KNQQILANAF NQAIGNITQA FGKVNDAIHQ TSQGLATVAK ALAKVQDVVN TQGQALSHLT
Por Resp C  KNQQILASAF NQAIGNITQS FGKVNDAIHQ TSRGLTTVAK ALAKVQDVVN TQGQALRHLT
OC43        QNQKLIANAF NNALYAIQEG FDATN----- ----------S ALVKIQAVVN ANAEALNNLL
BoCoV       QNQKLIANAF NNALDAIQEG FDATN----- ----------S ALVKIQAVVN ANAEALNNLL
MHV         ENQKMIASAF NNAIGAIQEG FAATN----- ----------S ALAKMQFVVN ANAEALNNLL
Rat CoV     ENQKMIASSF NNAIGAIQEG FDATN----- ----------S ALAKIQSVVN ANAEALNNLL
PHEV        QNQKLIASAF NNALDAIQEG FDATN----- ----------S ALVKIQAVVN ANAEALNNLL
AIBV        KNQEKIAASF NKAIGHMQEG FRSTS----- ----------L ALQQIQDVVS KQSAILTETM
SARS        ENQKQIANQF NKAISQIQES LTTTS----- ----------T ALGKLQDVVN QNAQALNTLV

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    1205       1215       1225       1235       1245       1255
EMCR S      SQLRHNFQAI SNSIHAIYDR LDSIQADQQV DRLITGRLAA LNAFVSQVLN KYTEVRGSRR
229E S      SQLRQNFQAI SSSIQAIYDR LDTIQADQQV DRLITGRLAA LNVFVSHTLT KYTEVRASRQ
PEDV        VQLQHNFQAI SSSIDDIYSR LDILSADVQV DRLITGRLSA LNAFVAQTLT KYTEVQASRK
TGEV        VQLQNNFQAI SSSISDIYNR LDELSADAQV DRLITGRLTA LNAFVSQTLT RQAEVRASRQ
CaCoV       VQLQNNFQAI SSSISDIYNR LDELSADAQV DRLITGRLTA LNAFVSQTLT RQAEVRASRQ
FeCoV       VQLQNNFQAI SSSISDIYNR LDELSADAQV DRLITGRLTA LNAFVSQTLT RQAEVRASRQ
Por Resp C  VQLQNNFQAI SSSISDIYNR LDELSADAQV DRLITGRLTA LNAFVSQTLT RQAEVRASRQ
OC43        QQLSNRFGAI SASLQEILSR LDALEAEAQI DRLINGRLTA LNAYVSQQLS DSTLVKFSAA
BoCoV       QQLSNRFGAI SSSLQEILSR LDALEAQAQI DRLINGRLTA LNVYVSQQLS DSTLVKFSAA
MHV         NQLSNRFGAI SASLQEILSR LDALEAQAQI DRLINGRLTA LNAYVSKQLS DMTLVKVSAA
Rat CoV     NQLSNRFGAI SASLQEILSR LDALEAQAQI DRLINGRLTA LNAYVSKQLS DMTLVKVSAA
PHEV        QQLSNRFGAI SASLQEILSR LDALEAKAQI DRLINGRLTA LNAYVSQQLS DSTLVKFSAA
AIBV        ASLNKNFGAI SSVIQEIYQQ FDAIQANAQV DRLITGRLSS LSVLASAKQA EYIRVSQQRE
SARS        KQLSSNFGAI SSVLNDILSR LDKVEAEVQI DRLITGRLQS LQTYVTQQLI RAAEIRASAN

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    1265       1275       1285       1295       1305       1315
EMCR S      LAQQKINECV KSQSNRYGFC G-NGTHIFSI VNSAPDGLLF LHTVLLPTDY KNVKAWSGIC
229E S      LAQQKVNECV KSQSKRYGFC G-NGTHIFSI VNAAPEGLVF LHTVLLPTQY KDVEAWSGLC
PEDV        LAQQKVNECV KSQSQRYGFC GGDGEHIFSL VQAAPQGLLF LHTVLVPGDF VNVLAIAGLC
TGEV        LAKDKVNECV RSQSQRFGFC G-NGTHLFSL ANAAPNGMIF FHTVLLPTAY ETVTAWPGIC
CaCoV       LAKDKVNECV RSQSQRFGFC G-NGTHLFSL ANAAPNGMIF FHTVLLPTAY ETVTAWPGIC
FeCoV       LAKDKVNECV RSQSQRFGFC G-NGTHLFSL ANAAPNGMIF FHTVLLPTAY ETVTAWSGIC
Por Resp C  LAKDKVNECV RSQSQRFGFC G-NGTHLFSL ANAAPNGMIF FHTVLLPTAY ETVTAWSGIC
OC43        QAMEKVNECV KSQSSRINFC G-NGNHIISL VQNAPYGLYF IHFSYVPTKY VTARVSPGLC
BoCoV       QAMEKVNECV KSQSSRINFC G-NGNHIISL VQNAPYGLYF IHFSYVPTKY VTAKVSPGLC
MHV         QAIEKVNECV KSQSSRINFC G-NGNHILSL VQNAPYGLYF IHFSYVPTSF TTANVSPGLC
Rat CoV     QAIEKVNECV KSQSPRINFC G-NGNHILSL VQNAPYGLYF IHFSYVPTSF TTANVSPGLC
PHEV        QAIEKVNECV KSQSSRINFC G-NGNHIISL VQNAPYGLYF IHFSYVPTKY VTAKVSPGLC
AIBV        LATQKINECV KSQSIRYSFC G-NGRHVLTI PQNAPNGIVF IHFSYTPDSF VNVTAIVGFC
SARS        LAATKMSECV LGQSKRVDFC G-KGYHLMSF PQAAPHGVVF LHVTYVPSQE RNFTTAPAIC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    1325       1335       1345       1355       1365       1375
EMCR S      VDG----IYG YVLRQPNLVL YS------DN GVFRVTSRVM FQPRLPVLSD FVQIYNCNVT
229E S      VDG----TNG YVLRQPNLAL YK------EG NYYRITSRIM FEPRIPTMAD FVQIENCNVT
PEDV        VNG----EIA LTLREPGLVL FTHELQTYTA TEYFVSSRRM FEPRKPTVSD FVQIESCVVT
TGEV        ASDG-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPRTM YQPRVATSSD FVQIEGCDVL
CaCoV       ASDG-SRTFG LVVEDVQLTL FRN-----LD EKFYLTPRTM YQPRVATSSD FVQIEGCDVL
FeCoV       ASDG-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPRTM YQPRVATSSD FVQIEGCDVL
Por Resp C  ALDV-DRTFG LVVKDVQLTL FRN-----LD DKFYLTPRTM YQPRVATSSD FVQIEGCDVL
```

```
OC43        IAG----DRG IAPKSGYFVN VN-------- NTWMYTGSGY YYPEPITENN VVVMSTCAVN
BoCoV       IAG----DRG IAPKSGYFVN VN-------- NTWMFTGSGY YYPEPITGNN VVVMSTCAVN
MHV         ISG----DRG LAPKAGYFVQ DD-------- GEWKFTGSNY YYPEPITDKN SVVMSSCAAN
Rat CoV     ISG----DRG LAPKAGYFVQ DH-------- GEWKFTGSNY YYPESITDKN SVVMSSCAVN
PHEV        IAG----DIG ISPKSGYFIN VN-------- NSWMFTGSSY YYPEPITQNN VVVMSTCAVN
AIBV        VKPANASQYA IVPANGRGIF IQVN------- GSYYITARDM YMPRAITAGD VVTLTSCQAN
SARS        HEG-----KA YFPREGVFVF NG-------- TSWFITQRNF FSPQIITTDN TFVSGNCDVV

                    1385       1395       1405       1415       1425       1435
EMCR S      FVNISRVELH TVIP-DYVDV NKTLQEFAQN L-PKYVKPNF DLTPFNLTYL NLSSELKQLE
229E S      FVNISRSELQ TIVP-EYIDV NKTLQELSYK L-PNYTVPDL VVEQYNQTIL NLTSEISTLE
PEDV        YVNLTSDQLP DVIP-DYIDV NKTLDEILAS L-PNRTGPSL PLDVFNATYL NLTGEIADLE
TGEV        FVNATVSDLP SIIP-DYIDI NQTVQDILEN FRPNWTVPEL TFDIFNATYL NLTGEIDDLE
CaCoV       FVNGTVIELP SIIP-DYIDI NQTVQDILEN FRPNWTVPEL PLDIFHATYL NLTGEINDLE
FeCoV       FVNATVIDLP SIIP-DYIDI NQTVQDILEN YRPNWTVPEF TLDIFNATYL NLTGEIDDLE
Por Resp C  FVNTTVSDLP SIIP-DYIDI NQTVQDILEN FRPNWTVPEL TLDVFNATYL NLTGEIDDLE
OC43        YTKAPYVMLN TSIP-NLPDF KEELDQWFKN QTSVAPDLSL DY--INVTFL DLQVEMN---
BoCoV       YTKAPDVMLN ISTP-NLHDF KEELDQWFKN QTSVAPDLSL DY--INVTFL DLQDEMN---
MHV         YTKAPEVFLN TSIP-NLPDF KEELDKWFKN QTSIAPDLSL DFEKLNVTLL DLTDEMN---
Rat CoV     YTKAPEVFLN TSIT-NLPDF KEELDKWFKN QTSIVPDLSF DIGKLNVTFL DLSYEMN---
PHEV        YTKAPDLMLN TSTP-NLPDF KEELYQWFKN QSSVAPDLSL DY--INVTFL DLQDEMN---
AIBV        YVSVNKTVIT TFVDNDDFDF NDELSKWWND T--KHELPDF DKFNYTVPIL DIDSEID---
SARS        IGIINNTVYD PLQP-ELDSF KEELDKYFKN HTSPDVDLGD ISG-INASVV NIQKEID---

                    1445       1455       1465       1475       1485       1495
EMCR S      AKTASLFQTT VELQGLIDQI NSTYVDLKLL NRFENYIKWP WWVWLIISVV FVVLLSLLVF
229E S      NKSAELNYTV QKLQTLIDNI NSTLVDLKWL NRVETYIKWP WWVWLCISVV LIFVVSMLLL
PEDV        QRSESLRNTT EELRSLINNI NNTLVDLEWL NRVETYIKWP WWVWLIIVIV LIFVVSLLVF
TGEV        FRSEKLHNTT VELAILIDNI NNTLVNLEWL NRIETYVKWP WYVWLLIGLV VIFCIPILLF
CaCoV       FRSEKLHNTT VELAILIDNI NNTLVNLEWL NRIETYVKWP WYVWLLIGLV VIFCIPILLF
FeCoV       FRSEKLHNTT VELAILIDNI NNTLVNLEWL NRIETYVKWP WYVWLLIGLV VVFCIPLLLF
Por Resp C  FRSEKLHNTT VELAILIDNI NNTVVNLEWL NRIETYVKWP WYVWLLIGLV VIFCIPLLLF
OC43        ---------- -RLQEAIKVL NQSYINLKDI GTYEYYVKWP WYVWLLICLA GVAMLVLLFF
BoCoV       ---------- -RLQEAIKVL NQSYINLKDI GTYEYYVKWP WYVWLLIGFA GVAMLVLLFF
MHV         ---------- -RIQDAIKKL NESYINLKDV GTYEMYVKWP WYVWLLIGLA GVAVCVLLFF
Rat CoV     ---------- -RIQDAIKNL NESYINLKEI GTYEMYVKWP WYVWLLIGLA GVAVCVLLFF
PHEV        ---------- -RLQEAIKVL NQSYINLKDI GTYEYYVKWP WYVWLLIGLA GVAMLVLLFF
AIBV        ---------- -RIQGVIQGL NDSLIDLEKL SILKTYIKWP WYVWLAIAFA TIIFILILGW
SARS        ---------- -RLNEVAKNL NESLIDLQEL GKYEQYIKWP WYVWLGFIAG LIAIVMVTIL

                    1505       1515       1525       1535       1545
EMCR S      CCLSTGCCGC CNCLTSSMRG CCDCGSTKLP YYEFEKVHVQ ----------
229E S      CCCSTGCCGF FSCFASSIRG CCES--TKLP YYDVEKIHIQ ----------
PEDV        CCISTGCCGC CGCCGACFSG CCRG-PRLQP YEAFEKVHVQ ----------
TGEV        CCCSTGCCGC IGCLGSCCHS ICSR-RQFEN YEPIEKVHVH ----------
CaCoV       CCCSTGCCGC IGCLGSCCHS ICSR-GQFES YEPIEKVHVH ----------
FeCoV       CCFSTGCCGC IGCLGSCCHS ICSR-RQFEN YEPIEKVHVH ----------
Por Resp C  CCCSTGCCGC IGCLGSCCHS IFSR-RQFEN YEPIEKVHVH ----------
OC43        ICCCTG-CG- -TSCFKKCGG CCDDYTGYQE LVIKT---SH DD--------
BoCoV       ICCCTG-CG- -TSCFKICGG CCDDYTGHQE LVIKT---SH DD--------
MHV         ICCCTG-CG- -SCCFKKCGN CCDECGGHQD SIVIHNISSH ED--------
Rat CoV     ICCCTG-CG- -SCCFKKCGN CCDEYGGRQA GIVIHNISSH ED--------
PHEV        ICCCTG-CG- -TSCFKKCGG CCDDYTGHQE FVIKT---SH DD--------
AIBV        VFFMTGCCGC CCGCFGIMPL MSKCGKKSSY YTTFDNDVVT EQYRPKKSV
SARS        LCCMTSCCS- -CLKGACSCG SCCKFDEDDS EPVLKGVKLH YT--------
```

f. Putative Orf 4a

```
                    5          15         25         35         45         55
EMCR 4a     MPFGGLFQLT LESTINKSVA NLKLPPHDVT VLRDNLKPVT TLSTITAYLL VSLFVTYFAL
229E 4a     MALG-LFTLQ LVSAVNQSLS NAKVSAEVSR QVIQDVKDGT VTFNLLAYTL MSLFVVYFAL

                    65         75         85         95         105        115
EMCR 4a     FKPLTARGRV ACFVLKLLTL SVYVPLLVLF GMYLDSFIIF FLRCCFDSYM LAIMPISNKN
229E 4a     FKARSHRGRA ALIVFKILIL FVYVPLLYWS QAYIYATLIA VILLG-RFFH TAWHCWLYKT

                    125        135        145        155        165        175
EMCR 4a     FSFVLFNVTK LCFVSGKCWY LEQSFYENRF AAIYGGDHYV VLGGETITFV SFDDLYVAIR
229E 4a     WDFIVFNVTT LCYAR----- ---------- ---------- ---------- ----------
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|.
                        185        195        205        215        225
EMCR 4a             GSCEKNLQLM RKVDLYNGAV IYIFAEEPVV GIVYSSQLYE DVPSIN
229E 4a             ---------- ---------- ---------- ---------- ------
```

g. Putative Orf 4ab

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         5         15         25         35         45         55
EMCR 4a             MPFGGLFQLT LESTINKSVA NLKLPPHDVT VLRDNLKPVT TLSTITAYLL VSLFVTYFAL
229E 4a             MALG-LFTLQ LVSAVNQSLS NAKVSAEVSR QVIQDVKDGT VTFNLLAYTL MSLFVVYFAL
229E 4b             ---------- ---------- ---------- ---------- ---------- ----------

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         65         75         85         95        105        115
EMCR 4a             FKPLTARGRV ACFVLKLLTL SVYVPLLVLF GMYLDSFIIF FLRCCFDSYM LAIMPISNKN
229E 4a             FKARSHRGRA ALIVFKILIL FVYVPLLYWS QAYIYATLIA VILLG-RFFH TAWHCWLYKT
229E 4b             ---------- ---------- ---------- ---------- ---------- ----------

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        125        135        145        155        165        175
EMCR 4a             FSFVLFNVTK LCFVSGKCWY LEQSFYENRF AAIYGGDHYV VLGGETITFV SFDDLYVAIR
229E 4a             WDFIVFNVTT LCYAR----- ---------- ---------- ---------- ----------
229E 4b             ---------- ----MQGKCW FLENKALKPF VCFYGGDQFL YIGDRIVSYF STNDLYVALR

                    ....|....| ....|....| ....|....| ....|....| ....|.
                        185        195        205        215        225
EMCR 4a             GSCEKNLQLM RKVDLYNGAV IYIFAEEPVV GIVYSSQLYE DVPSIN
229E 4a             ---------- ---------- ---------- ---------- ------
229E 4b             GRIDKDLSLS RKVELYNGEC VYLFCEHPAV GIVNTDFKLE IH....
```

h. Putative Orf E

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         5         15         25         35         45         55
EMCR E              ----MFLRLI DDNG-IVLNS ILWLLVMIFF F-VLAMTFIK LIQLCFTCHY FFSRTLYQP-
229E                ----MFLKLV DDHA-LVVNV LLWCVVLIVI L-LVCITIIK LIKLCFTCHM FCNRTVYGP-
PEDV                -----MLQLV NDNG-LVVNV ILWLFVLFFL L-IISITFVQ LVNLCFTCHR LCNSAVYTP-
TGEV                MTFPRALTVI DDNG-MVINI IFWFLLIIIL I-LLSIALLN IIKLCMVCCN LGRTVIIVP-
CaCoV               MTFPRALTVI DDNG-MVISI IFWFLLIIIL I-LFSIALLN IIKLCMVCCN LGRTVIIVP-
FeCoV               MTFPRAFTII DDHG-MVVSV FFWLLLIIIL I-LFSIALLN VIKLCMVCCN LGKTIIVLP-
Por Resp C          MTFPRALTVI DDNG-MVISI IFWFLLIIIL I-LLSIALLN IIKLCMVCCN LGRTVIIVP-
OC43                --MFMADAYL ADTV-WYVGQ IIFIVAICLL VTIVVVAFLA TFKLCIQLCG MCNTLVLSP-
BoCoV               --MFMADAYF ADTV-WYVGQ IIFIVAICLL VIIVVVAFLA TFKLCIQLCG MCNTLVLSP-
PHEV                --MFMADAYL ADTV-WYVGQ IIFIVAICLL VIIVVVAFLA TFKLCIQLCG MCNTLVLSP-
MHV                 ----MFNLFL TDTV-WYVGQ IIFIVAVCLM VTIIVVAFLA SIKLCIQLCG LCNTLLLSP-
Rat CoV             ----MFNLFL IDTV-WYVGQ IIFIVAVCLM VTIIVVAFLA SIKLCIQLCG LCNTLLLSP-
AIBV                --MNLLNKSL EENG-SFLTA LYIIVGFLAL Y-LLGRALQA FVQAADACCL FWYTWVVIPG
SARS                ----MYSFVS EETGTLIVNS VLLFLAFVVF L-LVTLAILT ALRLCAYCCN IVNVSLVKP-

                    ....|....| ....|....| ....|....| ....|....| ....|....| ..
                         65         75         85         95        105
EMCR E              ---VYKIFL- ---------- ----AYQDYM --QIAPV-PA EVLNV----- --
229E                ---IKNVYH- ---------- ----IYQSYM --HIDPF-PK RVIDF----- --
PEDV                ---IGRLYR- ---------- ----VYKSYM --RIDPL-PS TVIDV----- --
TGEV                ---AQHAYD- ---------- ----AYKNFM --RIKAYNPD GALLA----- --
CaCoV               ---ARHAYD- ---------- ----AYKNFM --QIRAYNPD EALLV----- --
FeCoV               ---ARHAYD- ---------- ----AYKTFM --QTKAYNPD EAFLV----- --
Por Resp C          ---VQHAYD- ---------- ----AYKNFM --RIKAYNPD GALLV----- --
OC43                ---SIYVENR GR-------- ----QFYEFYN --DVKPP-VL DVDDV----- --
BoCoV               ---SIYVFNR GR-------- ----QFYEFYN --DVKPP-VL DVDDV----- --
PHEV                ---SIYVFNR GR-------- ----QFYEFYN --DVKPP-VL DVDDV----- --
MHV                 ---SICVYNR SK-------- ----QLYKYYN E-EVRPP-PL EVDDIIIQTL --
Rat CoV             ---SIYVYNR SK-------- ----QLYKYYN E-EVRPP-PL EVDDIIIQTL --
AIBV                AKGTAFVYKY TYGRKLNNPE LEAVIVNEFP KNGWNNKNPA NFQDAQRDKL YS
SARS                ---TVYVYS- ---------- ----RVKNLN --SSEGV-PD LLV-------- --
```

i. Putative Orf M (Matrix protein)

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
```

```
              5          15         25         35         45         55
EMCR    ---------- ---------- ---------- ---------- ---------M SNSS------
229E    ---------- ---------- ---------- ---------- ---------M SNDN------
PEDV    ---------- ---------- ---------- ---------- ---------M SNGS------
TGEV    ---------- ---------- --------MK ILLILACVIA CACGERYCAM KSDTDLSCRN
CaCoV   ---------- ---------- -------MKK ILFLLACAIA CVYGERYCAM TESS-TSCRN
FeCoV   MHMMPIRPLC KPRHIIPTKH FWFELNKMKY ILLILACIIA CVYGERYCAM QDSG-LQCIN
PRCoV   ---------- ---------- --------MK ILLILACAIA CTCGERYCAM KDDTGLSCRN
OC43    ---------- ---------- ---------- ---------- ---------M SSKT------
PHEV    ---------- ---------- ---------- ---------- ---------M SSPT------
BoCoV   ---------- ---------- ---------- ---------- ---------M SSVT------
MHV     ---------- ---------- ---------- ---------- ---------M TSTTQ-----
RatSAV  ---------- ---------- ---------- ---------- ---------M SSTTP-----
AIBV    ---------- ---------- ---------- ---------- ---------M PNETN-----
SARS    ---------- ---------- ---------- ---------- ---------M ADNG------

        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
             65         75         85         95        105        115
EMCR    ---------V PLSEVYVHLR NWNFSWNLIL TVFIVVLQYG HYKYSRLLYG LKMSVLWCLW
229E    ---------C -TGDIVTHLK NWNFGWNVIL TIFIVILQFG HYKYSRLFYG LKMLVLWLLW
PEDV    ---------I PVDEVIEHLR NWNFTWNIIL TILLVVLQYG HYKYSVFLYG VKMAILWILW
TGEV    STASDCESCF NGGDLIWHLA NWNFSWSIIL IVFITVLQYG RPQFSWFVYG IKMLIMWLLW
CaCoV   STAGNCASCF ETGDLIWHLA NWNFSWSVIL IIFITVLQYG RPQFSWFVCG IKMLIMWLLW
FeCoV   GTNSRCQTCF ERGDLIWHLA NWNFSWSVIL IVFITVLQYG RPQFSWLVYG IKMLIMWLLW
PRCoV   GTASDCESCF NRGDLIWLLA NWNFSWSIIL IIFITVLQYG RPQFSWFVYG IKMLIMWLLW
OC43    --TPAPVYIW TADEAIKFLK EWNFSLGIIL LFITIILQFG YTSRSMFVYV IKMIILWLMW
PHEV    --TPVPVISW TADEAIKFLK EWNFSLGIIV LFITIILQFG YTSRSMFVYV IKMVILWLMW
BoCoV   --TPAPVYTW TADEAIKFLK EWNFSLGIIL LFITIILQFG YTSRSMFVYV IKMIILWLMW
MHV     --APQPVYQW TADEAIRFLK EWNFSLGIIL LFVTIILQFG YTSRSMFVYV VKMILLWLMW
RatSAV  --APQTVYQW TADVAVRFLK EWNFLLGIIL LFITIILQFG YTSRSMFIYV VKMIILWLMW
AIBV    -------CTL DFEQSVQLFK EYNLFITAFL LFLTIILQYG YATRSKVIYT LKMIVLWCFW
SARS    --------TI TVEELKQLLE QWNLVIGFLF LAWIMLLQFA YSNRNRFLYI IKLVFLWLLW

        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
            125        135        145        155        165        175
EMCR    PLVLALSIFD CFVNFNVD-W VFFGFSILMS IITLCLWVMY FVNSFRLWRR VKTFWAFNPE
229E    PLVLALSIFD TWANWDSN-W AFVAFSFFMA VSTLVMWVMY FANSFRLFRR ARTFWAWNPE
PEDV    PLVLALSLFD AWASFQVN-W VFFAFSILMA CITLMLWIMY FVNSIRLWRR THSWWSFNPE
TGEV    PVVLALTIFN AYSEYQVSRY VMFGFSIAGA IVTFVLWIMY FVRSIQLYRR TKSWWSFNPE
CaCoV   PIVLALTIFN AYLEYRVSRY VMFGFSVAGA TVTFILWIMY FVRSIQLYRR TKSWWSFNPE
FeCoV   PIVLALTIFN AYSEYQVSRY VMFGFSVAGA VVTFALWMMY FVRSVQLYRR TKSWWSFNPE
PRCoV   PIVLALTIFN AYSEYQVSRY VMFGFSIAGA IVTFVLWIMY FVRSIQLYRR TKSWWSFNPE
OC43    PLTIILTIFN --CVYALN-N VYLGLSIVFT IVAIIMWIVY FVNSIRLFIR TGSFWSFNPE
PHEV    PLTIILTIFN --CVYALN-N VYLGFSIVFT IVAIIMWVVY FVNSIRLFIR TGSWWSFNPE
BoCoV   PLTIILTIFN --CVYALN-N VYLGFSIVFT IVAIIMWIVY FVNSIRLFIR TGSWWSFNPE
MHV     PLTIVLCIFN --CVYALN-N VYLGFSIVFT IVSIIMWIMY FVNSIRLFIR TGSWWSFNPE
RatSAV  PLTIVLCIFN --CVYALN-N VYLGFSIVFT IVSIVMWIMY FVNSIRLFIR TGSWWSFNPE
AIBV    PLNIAVGVIS --CTYPPN-T GGLVAAIILT VFACLSFVGY WIQSIRLFKR CRSWWSFNPE
SARS    PVTLACFVLA --AVYRIN-W VTGGIAIAMA CIVGLMWLSY FVASFRLFAR TRSMWSFNPE

        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
            185        195        205        215        225        235
EMCR    TNAIISLQVY -GHNYYLPVM AAPTGVTLTL LSGVLLVDGH KIATRVQVGQ LPKYVIVATP
229E    VNAITVTTVL -GQTYYQPIQ QAPTGITVTL LSGVLYVDGH RLASGVQVHN LPEYMTVAVP
PEDV    TDALLTTSVM -GRQVCIPVL GAPTGVTLTL LSGTLLVEGY KVATGVQVSQ LPNFVTVAKA
TGEV    TKAILCVSAL -GRSYVLPLE GVPTGVTLTL LSGNLYAEGF KIAGGMNIDN LPKYVMVALP
CaCoV   TSAILCVSAL -GRSYVLPLE GVPTGVTLTL LSGNLCAEGF KIAGGMNIDN LPKYVMVALP
FeCoV   TNAILCVNAL -GRSYVLPLD GTPTGVTLTL LSGNLYAEGF KMAGGLTIEH LPKYVMIATP
PRCoV   TNAILCVSAL -GRSYVLPLE GVPTGVTLTL LSGNLYAEGF KIAGGMTIDN LPKYVMVALP
OC43    TNNLMCIDMK -GTMYVRPII EDYHTLTVTI IRGHLYIQGI KLGTGYSLAD LPAYMTVAK-
PHEV    TNNLMCIDMK -GRMYVRPII EDYHTLTATI IRGHLYIQGI KLGTGYSLSD LPAYVTVAK-
BoCoV   TNNLMCIDMK -GRMYVRPII EDYHTLTVTI IRGHLYMQGI KLGTGYSLSD LPAYVTVAK-
MHV     TNNLMCIDMK -GTVYVRPII EDYHTLTATI IRGHLYMQGV KLGTGFSLSD LPAYVTVAK-
RatSAV  TNNLMCIDVK -GTVYVRPII EDYHTLTATN VRGHLYMQGV KLGTGFSLSD LPAYVTVAK-
AIBV    SNAVGSILLT NGQQCNFAIE SVPMVLSPII KNGVLYCEGQ WLAK-CEPDH LPKDIFVCTP
SARS    TNILLNVPLR -GTIVTRPLM ESELVIGAVI IRGHLRMAGH PLGR-CDIKD LPKEITVAT-

        ....|....| ....|....| ....|....| ....|....| ....|....| ..
            245        255        265        275        285
EMCR    STTIVCDRVG RSVNETSQTG WAFYVRAKHG DFSGVASQEG VLSEREKLLH LI
229E    STTIIYSRVG RSVNSQNSTG WVFYVRVKHG DFSAVSSPMS NMTENERLLH FF
PEDV    TTTIVYGRVG RSVNASSGTG WAFYVRSKHG DYSAVSNPSA VLTDSEKVLH LV
TGEV    SRTIVYTLVG KKLKASSATG WAYYVKSKAG DYSTEAR-TD NLSEQEKLLH MV
CaCoV   VRTIVYTLVG KKLKASSATG WAYYVKSKAG DYSTDAR-TD NLSEHEKLLH MV
FeCoV   SRTIVYTLVG KQLKATTATG WAYYVKSKAG DYSTEAR-TD NLSEHEKLLH MV
PRCoV   SRTIVYTLVG KKLKASSATG WAYYVKSKAG DYSTEAR-TD NLSEQEKLLH MV
OC43    VTHLCTYKRG FLDRISDTSG FAVYVKSKVG NYRLPSTQKG SGMDTALLRN NI
PHEV    VTHLCTYKRG FLDRIGDTSG FAVYVKSKVG NYRLPSTHKG SGMDTALLRN NI
BoCoV   VSHLLTYKRG FLDKIGDTSG FAVYVKSKVG NYRLPSTQKG SGMDTALLRN NI
MHV     VSHLCTYKRA FLDKVDGVSG FAVYVKSKVG NYRLPSN-KP SGMDTALLR- -I
```

```
RatSAV      VSHLCTYKRA FLDKVDGVSG FAVYVKSKVG NYRLPSN-KP SGADTALLR- -I
AIBV        DRRNIYRMVQ KYTGDQSGNK KRFATFVYAK QSVDTGELES VATGGSSLYT --
SARS        SRTLSYYKLG ASQRVGTDSG FAAYNRYRIG NYKLNTDHAG SNDNIALLVQ --
```

j, Putative Orf N (Nucleoprotein)

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     5         15         25         35         45         55
EMCR        ---------- --MAS----- ----VN---W ADDR------ --AARKKF-- ----------
229E        ---------- --MAT----- ----VK---W ADASEPQ--- --RGRQGR-- ----------
PEDV        ---------- --MAS----- ----VS---F QDRG------ --RKR----- ----------
TGEV        ---------- --MANQGQR- ----VS---W GDESTKT--- --RGRSNSRG RKNNN-----
FeCoV       ---------- --MATQGQR- ----VN---W GDEPSKR--- --RGRSNSRG RKNND-----
PRCoV       ---------- --MANQGQR- ----VS---W GDESTKI--- --RGRSNSRG RKINN-----
CaCoV       ---------- --MASQGQR- ----VS---W GDESTKR--- --RGRSNSRG RKNND-----
RSDACoV     MSFVPGQENA GSRSSSGNRA GNGILKKTTW ADQTERGQNN GNRGRRNQPK QTATTQ-PNT
MHV         MSFVPGQENA GSRSSSGNRA GNGILKKTTW ADQTERG--- -NRGRRNHPK QTATTQ-PNA
PHEV        MSFTPGKQSS -SRASSGNRS GNGILK---W ADQSDQSRNV QTRGRRVQSK QTATSQQPSG
OC43        MSFTPGKQSS -SRASSGNRS GNGILK---W ADQSDQFRNV QTRGRRAQPK QTATSQQPSG
BoCoV       MSFTPGKQSS -SRASFGNRS GNGILK---W ADQSDQSRNV QTRGRRAQPK QTATSQLPSG
SARS        ---------- --MSDNGPQS NQRSAPRITF GGPTDSTDNN QNGGRNGARP KQRRPQ----
AIBV        ---------- --MASG---- -----K---A AGKTDAPAPV IKLGGPKPPK VGSS------

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     65         75         85         95        105        115
EMCR        ----PPPSFY MPLLVSSDKA PYRVIPRNLV PIGKGNK-DE QIGYWNVQER --WRMRRGQR
229E        ----IPYSLY SPLLVDSE-Q PWKVIPRNLV PINKKDK-NK LIGYWNVQKR --FRTRKGKR
PEDV        ----VPLSLY APLRVTNDKP LSKVLANNAV PTNKGNK-DQ QIGYWNEQIR --WRMRRGER
TGEV        ----IPLSFF NPITLQQGSK FWNLCPRDFV PKGIGNR-DQ QIGYWNRQTR --YRMVKGQR
FeCoV       ----IPLSFY NPITLEQGSK FWNLCPRDLV PKGIGNK-DQ QIGYWNRQIR --YRIVKGQR
PRCoV       ----IPLSFF NPITLQQGAK FWNSCPRDFV PKGIGNR-DQ QIGYWNRQTR --YRMVKGQR
CaCoV       ----IPLSFF NPITLEQGSK FWDLCPRDFV PKGIGNK-DQ QIGYWNRQTR --YRMVKGRR
RSDACoV     GSVVPHYSWF SGITQFQKGK EFQFAGGQGV PIANGIPPSE QKGYWYRHNR RSFKTPDGQQ
MHV         GSVVPHYSWF SGITQFQKGK EFQFAQGQGV PIASGIPASE QKGYWYRHNR RSFKTPDGQH
PHEV        GTVVPYYSWF SGITQFQKGK EFEFAEGQGV PIAPGVPSTE AKGYWYRHNR RSFKTADGNQ
OC43        GNVVPYYSWF SGITQFQKGK EFEFVEGQGV PIAPGVPATE AKGYWYRHNR RSFKTADGNQ
BoCoV       GNVVPYYSWF SGITQFQKGK EFEFAEGQGV PIAPGVPATE AKGYWYRHNR RSFKTADGNQ
SARS        GLPNNTASWF TALTQHGK-E ELRFFPRGQGV PINTNSGPDD QIGYYRRATR R-VRGGDGKM
AIBV        ----GNASWF QAIKAKKLNT PPPKFEGSGV PDNENIKPSQ QHGYWRRQAR --FKPGKGGR

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    125        135        145        155        165        175
EMCR        VDLPPKVHFY YLGTGPHKDL KFRQRSDGVV WVAKEGAKTV NTSLGNRK-- RNQKPLEPKF
229E        VDLSPKLHFY YLGTGPHKDA KFRERVEGVV WVAVDGAKTE PTGYGVRR-- KNSEPEIPHF
PEDV        IEQPSNWHFY YLGTGPHGDL RYRTRTEGVF WVAKEGAKTE PTNLGVRK-- ASEKPIIPKF
TGEV        KELPERWFFY YLGTGPHADA KFKDKLDGVV WVAKDGAMNK PTTLGSRG-- ANNESKALKF
FeCoV       KELAERWFFY FLGTGPHADA KFKDKIDGVF WVARDGAMNK PTTLGTRG-- TNNESKPLRF
PRCoV       KELPERWFFY YLGTGPHADA KFKDKLDGVV WVAKDGAMNK PTTLGTRG-- TNNESKPLRF
CaCoV       KNLPEKWFFY YLGTGPHADA KFKQKLDGVV WVARGDSMTK PTTLGTRG-- TNNESKALKF
RSDACoV     KQLLPRWYFY YLGTGPHAGA SFGDSIEGVF WVANSQADTN TSADIVERDP SSHEAIPTRF
MHV         KQLLPRWYFY YLGTGPHAGA EYGDDIEGVV WVASQQADTK TTADVVERDP SSHEAIPTRF
PHEV        RQLLPRWYFY YLGTGPHAKD QYGTDIDGVF WVASNQADIN TPADIVDRDP SSDEAIPTRF
OC43        RQLLPRWYFY YLGTGPHAKD QYGTDIDGVY WVASNQADVN TPADIVDRDP SSDEAIPTRF
BoCoV       RQLLPRWYFY YLGTGPHAKD QYGTDIDGVF WVASNQADVN TPADILDRDP SSDEAIPTRF
SARS        KELSPRWYFY YLGTGPEASL PYGANKEGIV WVATEGALNT PKDHIGTRNP NNNAATVLQL
AIBV        KPVPDAWYFY YTGTGPAADL NWGDTQDGIV WVAAKGADTK SRSNQGTRDP DKFDQYPLRF

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    185        195        205        215        225        235
EMCR        SIALPPELSV VEFEDRSNNS SRASSRSSTR ---------- ------NNSR DS--------
229E        NQKLPNGVTV VEEPD----- SRAPSRSQSR ---------- ------SQSR ----------
PEDV        SQQLPSVVEI VEPNTPP--A SRANSRSRSR GNGNNRSRSP SNNRGNNQSR GNSQNRGNNQ
TGEV        DGKVPGEFQL EVNQS----- -RDNSRSRSQ ---------- ---------- ----------
FeCoV       DGKIPPQFQL EVNRS----- -RNNSRSGSQ ---------- ---------- ----------
PRCoV       DGKVPGEFQL EVNQS----- -RDNSRSRSQ ---------- ---------- ----------
CaCoV       DVKVPSEFHL EVNQL----- -RDNSRSRSQ ---------- ---------- ----------
RSDACoV     APGTVLPQGF YVEGS----- GRSAPASRSG ---------- ---------- ----------
MHV         APGTVLPQGF YVEGS----- GRSAPASRSG ---------- ---------- ----------
PHEV        PPGTVLPQGY YIEGS----- GRSAPNSRST ---------- ---------- ----------
OC43        PPGTVLPQGY YIEGS----- GRSAPNSRST ---------- ---------- ----------
BoCoV       PPGTVLPQGY YIEGS----- GRSAPNSRST ---------- ---------- ----------
SARS        PQGTTLPKGF YAEGS----- -RGGSQASSR ---------- ---------- ----------
AIBV        SDGGPDGNFR WDFIP----- --LN-RGRSG ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    245        255        265        275        285        295
```

```
                EMCR    SRSTSRQ--- ------QSR- TRSDSNQS-- -----S-SDL VAAVTLALKN LGFDN--QSK
                229E    GRGESKP--- ------QSRN PSSDRNHN-- -----SQDDI MKAVAAALKS LGFDKP-QEK
                PEDV    GRGASQNRGG NNNNNNKSRN QSNNRNQSND RGGVTSRDDL VAAVKDALKS LGIGEN-PDR
                TGEV    SRSRSRNR-- -----SQSRG RQQFNNKK-- ------DDSV EQAVLAALKK LGVDTE-KQQ
                FeCoV   SRSVSRNR-- -----SQSRG RHHSNNQ--- ------NNNV EDTIVAVLEK LGV-TD-KQ-
                PRCoV   SRSRSRNR-- -----SQSRG RQQSNNKK-- ------DDSV EQAVLAALKK LGVYTE-KQQ
                CaCoV   SRSQSRNR-- -----SQSRG RQLSNNKK-- ------DDNV EQAVLAALKK LGVDTE-KQQ
                RSDACoV SRSQSRGP-- ------NNRA RSSSNQRQ-- -----PASTV KPDMAEEIAA LVLAN---LG
                MHV     SRSQSRGP-- ------NNRA RSSSNQRQ-- -----PASAV KPDMAEEIAA LVLAK---LG
                PHEV    SRAPNRAPS- -----AGSRS RANSGNRT-- -----STPGV TPDMADQIAS LVLAK---LG
                OC43    SRTSSRASS- -----AGSRS RANSGNRT-- -----PTSGV TPDMADQIAS LVLAK---LG
                BoCoV   SRASSRASS- -----AGSRS RANSGNRT-- -----PTSGV TPDMADQIAS LVLAK---LG
                SARS    SSSRSRGN-- -----SRNST PGSSRGNS-- -----PARMA SGGGETALAL LLLDRLNQLE
                AIBV    -RSTAASS-- ----AAASRA PSREGSRG-- -----RRSDS GDDLIARAAK IIQDQ-----

                        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          305       315       325       335       345       355

                EMCR    SPSSSGTSTP K-------K- ---PNKPLSQ PRADKPS--- -QLKKPRWKR VPTR--EENV
                229E    DKKSAKTGTP KPSRNQSPAS SQTSAKSLAR SQSSETKEQK HEMQKPRWKR QPNDDVTSNV
                PEDV    HKQQQKPKQE K-SDN---SG KNTPKKNKSR ATSKERD--- -LKDIPEWRR IPKG--ENSV
                TGEV    QRSRSKSKER S---------- -----NSKTR DTTPKNE--- ---NKHTWKR TAGK---GDV
                FeCoV   -RSRSKPRER S---------- -----DSKPR DTTPKNA--- ---NKHTWKK TAGK---GDV
                PRCoV   QRSRSKSKER S---------- -----NSKTR DTTPKNE--- ---NKHTWKR TAGK---GDV
                CaCoV   -RSRSKSKER S---------- -----SSKTR DTTPKNE--- ---NKHTWKR TAGK---GDV
                RSDACoV -KDAGQPKQV T---------- -----KQSAK EVRQKIL--- ---NKPRQKR TPNK--QCPV
                MHV     -KDAGQPKQV T---------- -----KQSAK EVRQKIL--- ---TKPRQKR TPNK--QCPV
                PHEV    -KDATKPQQV T---------- -----KQTAK EVRQKIL--- ---NKPRQKR SPNK--QCTV
                OC43    -KDATKPQQV T---------- -----KHTAK EVRQKIL--- ---NKPRQKR SPNK--QCTV
                BoCoV   -KDATKPQQV T---------- -----KQTAK EIRQKIL--- ---NKPRQKR SPNK--QCTV
                SARS    SKVSGKGQQQ Q---------- ----GQTVTK KSAAEAS--- ---KKPRQKR TATK--QYNV
                AIBV    ---QKKGSRI T---------- -----KAKAD EMAHRRY--- ------CKRT IPPN---YRV

                        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          365       375       385       395       405       415

                EMCR    IQCFGPRDFN H---NMGDSD LVQNGVDAKG FPQLAELIPN QAALFFDSEV STDEVG----
                229E    TQCFGPRDLD H---NFGSAG VVANGVKAKG YPQFAELVPS TAAMLFDSHI VSKESG-----
                PEDV    AACFGPRGGF K---NFGDAE FVEKGVDASG YAQIASLAPN VAALLFGGNV AVRELA----
                TGEV    TRFYGARSSS A---NFGDTD LVANGSSAKH YPQLAECVPS VSSILFGSYW TSKEDG----
                FeCoV   TTFYGARSSS A---NFGDSD LVANGNAAKC YPQIAECVPS VSSIIFGSQW SAEEAG----
                PRCoV   TRFYGARSSS A---NFGDSD LVANGSSAKH YPQLAECVPS VSSILFGSYW TSKEDG----
                CaCoV   TKFYGARSSS A---NFGDSD LVANGNGAKH YPQLAECVPS VSSILFGSHW TAKEDG----
                RSDACoV QQCFGKRGPN Q---NFGGPE MLKLGTSDPQ FPILAELAPT PGAFFFGSKL ELVKKN--SG
                MHV     QQCFGKRGPN Q---NFGGSE MLKLGTSDPQ FPILAELAPT PSAFFFGSKL ELVKKN--SG
                PHEV    QQCFGKRGPN Q---NFGGGE MLKLGTSDPQ FPILAELAPT AGAFFFGSRL ELAKVQNLSG
                OC43    QQCFGKRGPN Q---NFGGGE MLKLGTSDPQ FPILAELAPT AGAFFFGSRL ELAKVQNLSG
                BoCoV   QQCFGKRGPN Q---NFGGGE MLKLGTSDPQ FPILAELAPT AGAFFFGSRL ELAKVQNLSG
                SARS    TQAFGRRGPE QTQGNFGDQD LIRQGTDYKH WPQIAQFAPS ASAFFGMSRI GMEVTP----
                AIBV    DQVFGPRTKG K-EGNFGDDK MNEEGIKDGR VTAMLNLVPS SHACLFGSRV TPKLQL----

                        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          425       435       445       455       465       475

                EMCR    --------DNV QITYT----Y KMLVAKDNKN LPKFIEQISA FTKPS----- SIKEMQSQSS
                229E    --------NTV VLTFT----T RVTVPKDHPH LGKFLEELNA FTR-------- EMQQHPLLNP
                PEDV    --------DSY EITYN----Y KMTVPKSDPN VELLVSQVDA FKTGN----- AKLQRKKEKK
                TGEV    --------DQI EVTFT----H KYHLPKDDPK TGQFLQQINA YARPS----- EVAKEQRKRK
                FeCoV   --------DQV KVTLT----H TYYLPKDDAK TSQFLEQIDA YKRPS----- EVAKDQRQRR
                PRCoV   --------DQI EVTFT----H KYHLPKDHPK TEQFLQQINA YASPS----- ELAKEQRKRK
                CaCoV   --------DQI EVTFT----H KYHLPKDDPK TGQFLQQINA YARPS----- EVAKEQRQRK
                RSDACoV GVDEPTKDVY ELQYSGAVRF DSTLPGFETI MKVLNENLNA YQNQA----- GGADVVSPKP
                MHV     GADEPTKDVY ELQYSGAIRF DSTLPGFETI MKVLTENLNA YQDQA----- GSVDLVSPKP
                PHEV    NPDEPQKDVY ELRYNGAIRF DSTLSGFETI MKVLNQNLNA YQHQE----- DGMMNISPKP
                OC43    NPDEPQKDVY ELRYNGAIRF DSTLSGFETI MKVLNENLNA YQQQ------ DGMMNMSPKP
                BoCoV   NLDEPQKDVY ELRYNGAIRF DSTLSGFETI MKVLNENLNA YQQQ------ DGMMNMSPKP
                SARS    --------SGT WLTYHGAIKL DDKDPQFKDN VILLNKHIDA YKTFP----- PTEPKKDKKK
                AIBV    --------DGL HLRFEFTTVV PCDDPQFDNY VKICDQCVDG VGTRPKDDEP KPKSRSSSRP

                        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          485       495       505       515       525       535

                EMCR    HVAQNTVLN- ---------- -ASIPES--- -KPLADDDSA IIEIVNEVLH ----------
                229E    SALEFNPSQ- ---------- -TSPATA--- -EPVRDEVSI ETDIIDEVN- ----------
                PEDV    NKRETTLQQH EEAIYDDVGA PSDVTHANLE WDTAVDGGDT AVEIINEIFD TGN-------
                TGEV    SRSKSAERS- ---------- EQDVVPDALI ENYTDVFDDT QVEIIDEVTN ----------
                FeCoV   SRSKSADK-- ---------- KPEELSVTLV EAYTDVFDDT QVEMIDEVTN ----------
                PRCoV   SRSKSAERS- ---------- EQEVVPDSLI ENYTDVFDDT QVEMIDEVTN ----------
                CaCoV   ARSKSVERV- ---------- EQEVVPDALT ENYTDVFDDT QVEIIDEVTN ----------
                RSDACoV QRKRGTKQT- -AQKEELDSI SVAKPKSAVQ RNVSRELTPE DRSLLAQILD DGVVPDGLDD
                MHV     PRRGRRQAQ- -EKKDEVDNV SVAKPKSLVQ RNVSRELTPE DRSLLAQILD DGVVPDGLED
                PHEV    QRQRGQKN-- --GQVENDNV SVAAPKSRVQ QNKSRELTAE DISLLKKMDE P------YTED
                OC43    QRQRGHKN-- --GQGENDNI SVAVPKSRVQ QNKSRELTAE DISLLKKMDE P------YTED
                BoCoV   QRQRGQKN-- --GQGENDNI SVAAPKSRVQ QNKSRELTAE DISLLKKMDE P------YTED
```

```
SARS        KTDEAQPLP- ---------- QRQKKQPTVT LLPAADMDDF SRQLQNSMSG ASADSTQA--
AIBV        ATRGNSPAPR QQRPKKEKKL KKQDDEADKA LTSDEERNNA QLEFYDEPKV INWGDAALGE

            ....

EMCR        ----
229E        ----
PEDV        ----
TGEV        ----
FeCoV       ----
PRCoV       ----
CaCoV       ----
RSDACoV     -SNV
MHV         DSNV
PHEV        TSEI
OC43        TSEI
BoCoV       TSEI
SARS        ----
AIBV        NEL-
```

k. 5'untranslated region (genomic sequence)

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         5          15          25          35          45          55
EMCR5'UTR           ----------  ----------  ---AGATAGA GAATTTTCTT ATTTAGACTT TGTGTCTACT
229E5'UTR           ACTTAAGTAC CTTATCTATC TACAGATAGA AAAGTTGCTT -TTTAGACTT TGTGTCTACT

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         65          75          85          95          105         115
EMCR5'UTR           CCTCTCAACT AAACGAAATT TTT-CTAGTG CTGTCATTTG TTATG--GCA GTCCTAGTGT
229E5'UTR           TTTCTCAACT AAACGAAATT TTTGCTATGG CCGGCATCTT TGATGCTGGA GTCGTAGTGT

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         125         135         145         155         165         175
EMCR5'UTR           AATTGAAATT TCGTCAAGTT TGTAA-ACTG GTTAGGCAAG TGTTGTATTT TCTGTGTTTA
229E5'UTR           AATTGAAATT TCATTTGGGT TGCAACAGTT TGGAAGCAAG TGCTGTGTGT CCTA-GTCTA

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         185         195         205         215         225         235
EMCR5'UTR           AGCACTGGTG GTTCTGTC-C ACTAGTGCAC AC-ATTGATA CTTAAGT-GG TGTTCTGTCA
229E5'UTR           AGGGTTTCGT GTTCCGTCAC GAGATTCCAT TCTACAAACG CCTTACTCGA GGTTCCGTCT

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....
                         245         255         265         275         285
EMCR5'UTR           CTGCTTATTG TGGAAGCAAC GTTCTGTCGT TGTGGAAACC AATAACTGCT AACC
229E5'UTR           CGTGTTTGTG TGGAAGCAAA GTTCTGTCTT TGTGGAAACC AGTAACTGTT CCTA
```

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/NL04/000805

International filing date: 18 November 2004 (18.11.2004)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 03078772.5
Filing date: 01 December 2003 (01.12.2003)

Date of receipt at the International Bureau: 31 January 2005 (31.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)